Speech-Based Real-Time Presentation Tracking
Using Semantic Matching

A dissertation presented
by

Reza Asadi

to the Faculty of the Graduate School
of the College of Computer and Information Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Northeastern University
Boston, Massachusetts

December 2017
THESIS TITLE: Speech-Based Real-Time Presentation Tracking Using Semantic matching

AUTHOR: Reza Asadi

Ph.D. Thesis Approved to complete all degree requirements for the Ph.D. Degree in Computer Science.

Hamid Telle
Thesis Advisor

12/19/2017
Date

Rohit Garg
Thesis Reader

12/19/2017
Date

12/19/2017
Date

1/7/2018
Date

Graduate School Approval:

Rajiv
Director, Graduate School

1/9/2018
Date

Copy Received in Graduate School Office:

Megan Gravy
Recipient's Signature

11/9/2018
Date

Distribution: Once completed, this form should be scanned and attached to the front of the electronic dissertation document (page 1). An electronic version of the document can then be uploaded to the Northeastern University-UMI website.
Abstract

Oral presentations are an essential yet challenging aspect of academic and professional life. To date, many commercial and research products have been developed to provide support for the authoring, rehearsal and delivery of presentations. However, little work has been conducted to provide real-time tracking of a speaker’s presentation relative to their supporting media. Given the content of presentation slides and speaking notes, a presentation tracking system uses automatic speech recognition and semantic text retrieval techniques to track content coverage by the speaker. This can help speakers ensure that they cover their planned content, and enable various real-time presentation support technologies, such as automatic content recommendation during presentation delivery. Presentation tracking is, however, a complex task due to the inaccuracy of current speech recognition systems and the fact that speakers rarely follow their presentation notes exactly.

In this dissertation, I present a novel framework for real-time tracking of presentations at the sub-slide level, as well as global presentation tracking through a slide deck that allows for more speaker flexibility in choosing which slides to present. Tracking is performed by semantic matching of the confusion network results from an automatic speech recognition system against content keywords in slide text and notes. The keywords are selected and weighted based on word specificity and semantic similarity measures. My evaluation studies show that using confusion networks results in a more robust speech recognition system, while semantic matching reduces the reliance on the exact notes, and keyword weighting improves the accuracy of the tracking system.

The thesis of this dissertation is that utilizing presentation tracking in presentation assistance applications can improve the experience and the performance of presenters during rehearsal and delivery of presentations and QA sessions. To demonstrate the effectiveness of presentation tracking, I integrated the tracking framework into three different systems:
First, I designed and evaluated IntelliPrompter, a speech-based note display system that automatically tracks a presenter’s coverage of each slide’s content and dynamically adjusts the note display interface to highlight the most likely next topic to present. I investigated two modalities for dynamic note display: a conventional computer screen and a head-mounted display (Google Glass). The design of the system was informed by findings from 36 interviews with presenters and analysis of a presentation note corpus. In a within-subjects study comparing the dynamic screen-based and Google Glass note display interfaces with a static note system, presenters and independent judges expressed a strong preference for the dynamic screen-based system.

Second, I integrated the tracking framework into Quester, a system that enables fast access to relevant presentation content during a question answering session and supports nonlinear presentations led by the speaker. Given the slides’ content and notes, the system dynamically ranks presentation slides based on semantic closeness to spoken utterances, displays the most related slides, and highlights the corresponding content keywords in slide notes. The design of this system was informed by findings from interviews with expert presenters and analysis of recordings of lectures and conference presentations. In a within-subjects study comparing the dynamic support system with a static slide navigation system during a question answering session, presenters expressed a strong preference for the dynamic support system and answered the questions more efficiently using this system.

Finally, I evaluated the effectiveness of integrating the tracking framework into RoboCOP, a robot that acts as a public speaking coach to provide spoken feedback during presentation rehearsals. Results of a within-subjects evaluation study comparing RoboCOP with existing rehearsal practices showed that this system led to improved presentation quality.
Acknowledgment

First and foremost, I would like to thank my advisors, Harriet Fell and Timothy Bickmore, for their invaluable support and advice during my PhD. I would also like to thank Darren Edge and Lu Wang for being on my PhD committee and for the insightful feedback they have given me on my work.

I want to thank the many friends and labmates, who have assisted me throughout my endeavors. I am deeply indebted to Ha Trinh, for her help and guidance when I needed them the most. I would like to give a special thanks to Lazlo Ring, Dina Utami, Juan Fernandez, Ameneh Shamekhi, Zessie Zhang, Shuo Zhou, Stefan Olafsson, and Everlyne Kimani for being not only great lab mates, but awesome friends throughout my academic journey.

Lastly, I reserve the most special thanks for my family, who have given me endless love and support over the past years. It was thanks to my parents and grandparents’ encouragement that I decided to start my PhD journey. Thank you for everything.
Contents

ABSTRACT ................................................................................................................................. I

ACKNOWLEDGMENT ............................................................................................................... III

CONTENTS ................................................................................................................................ IV

LIST OF TABLES ......................................................................................................................... X

LIST OF FIGURES ..................................................................................................................... XI

INTRODUCTION ....................................................................................................................... 1

1.1 CURRENT SLIDEWARE TECHNOLOGY ............................................................................... 2

1.2 PROBLEMS OF CURRENT SLIDEWARE .............................................................................. 2

1.3 PRESENTATION TRACKING AND POTENTIAL APPLICATIONS ...................................... 4

1.4 CONTRIBUTIONS ................................................................................................................ 6

1.5 DISSERTATION STRUCTURE ............................................................................................... 7

BACKGROUND AND RELATED WORK ..................................................................................... 9

2.1 PRESENTATION TYPOLOGY ................................................................................................. 9

2.2 PRESENTATION ASSISTANCE TECHNOLOGIES ............................................................... 10

   2.2.1 Presentation Authoring Support .................................................................................. 10

   2.2.2 Presentation Rehearsal Support ................................................................................. 11

   2.2.3 Presentation Delivery Support .................................................................................. 13

   2.2.4 Non-Linear Presentation Support ............................................................................. 14

   2.2.5 Presentation Tracking .............................................................................................. 15

2.3 SPEECH RECOGNITION ..................................................................................................... 16

   2.3.1 Acoustic and Language Models ............................................................................... 17

   2.3.2 Confusion Networks ................................................................................................ 19

2.4 SEMANTIC TEXT RETRIEVAL ............................................................................................ 20

   2.4.1 Term Weighting ......................................................................................................... 20
2.4.2 Semantic Similarity .................................................................................................................. 21
2.5 CONCLUSION ............................................................................................................................ 23

EXPLORATORY STUDIES .................................................................................................................. 24

3.1 EXPLORATORY STUDY ON NOTES FORMS AND USAGE ....................................................... 24
  3.1.1 Current Practices in Presentation Note Taking ........................................................................ 25
  3.1.2 Application of Notes During Presentation Delivery ................................................................. 25
  3.1.3 Best Practices and Requested Features ................................................................................... 26
  3.1.4 Analyzing a Notes Corpus ...................................................................................................... 27
3.2 EXPLORATORY STUDY ON DYNAMIC PRESENTATIONS ...................................................... 28
  3.2.1 Need for Dynamic Presentations ............................................................................................ 28
  3.2.2 Preparing for Unscripted Interactions ..................................................................................... 29
  3.2.3 Nonlinear Slide Navigation ..................................................................................................... 30
3.3 IMPLICATIONS FOR DESIGN .................................................................................................. 31
3.4 CONCLUSION ............................................................................................................................. 32

TRACKING FRAMEWORK ............................................................................................................ 33

4.1 REQUIREMENT ANALYSIS ....................................................................................................... 33
  4.1.1 General Requirements ............................................................................................................. 33
  4.1.2 Prototype Development .......................................................................................................... 34
  4.1.3 Prototype Evaluation .............................................................................................................. 35
  4.1.4 Summary ................................................................................................................................ 39
4.2 DESIGN OF THE FRAMEWORK ................................................................................................. 39
  4.2.1 Slide Content Processing ........................................................................................................ 40
  4.2.2 Real-Time Content Matching ................................................................................................ 44
4.3 IMPLEMENTATION AND EVALUATION OF A TRACKING SYSTEM ......................................... 46
  4.3.1 Implementation Details .......................................................................................................... 46
  4.3.2 Evaluation Experiments ......................................................................................................... 47
4.3.3 Evaluation Results ........................................................................................................... 49
4.4 CONCLUSION ...................................................................................................................... 50

INTELLIPROMPTER SYSTEM ................................................................................................. 52

5.1 DESIGN OF THE NOTE DISPLAY SYSTEMS ................................................................... 53
  5.1.1 Authoring Environment ............................................................................................... 53
  5.1.2 Presentation Mode ........................................................................................................ 54

5.2 EVALUATION OF NOTE DISPLAY SYSTEMS ................................................................. 55
  5.2.1 Note Display Interfaces ............................................................................................... 56
  5.2.2 Procedure .................................................................................................................... 56
  5.2.3 Presenter Participants .................................................................................................. 58
  5.2.4 Presenter Measures ..................................................................................................... 58
  5.2.5 Presenter Quantitative Results ..................................................................................... 59
  5.2.6 Judge Evaluation of Presentation Quality ..................................................................... 62
  5.2.7 Objective Measures .................................................................................................... 63
  5.2.8 Qualitative Findings .................................................................................................... 66

5.3 CONCLUSION ..................................................................................................................... 68

QUESTER SYSTEM .................................................................................................................. 70

6.1 QUESTION ANSWERING SESSIONS ................................................................................. 71
  6.1.1 Guidelines ................................................................................................................... 71
  6.1.2 Analysis of Question Answering Sessions .................................................................... 71

6.2 DESIGN OF QUESTER ........................................................................................................ 72
  6.2.1 Note Authoring Environment ...................................................................................... 73
  6.2.2 Presentation Mode ....................................................................................................... 73
  6.2.3 Semantic Match Scoring .............................................................................................. 75
  6.2.4 Recency Model ........................................................................................................... 75
  6.2.5 Corpus of Presentations and Questions ....................................................................... 76
6.2.6 Model Evaluation ............................................................................................................... 78

6.3 USER EVALUATION OF QUESTER ..................................................................................... 81

6.3.1 Procedure ......................................................................................................................... 81

6.3.2 Participants ......................................................................................................................... 83

6.3.3 Self-report Measures of Presenter Experience ................................................................. 83

6.3.4 Self-perceived Rating of QA Performance ......................................................................... 85

6.3.5 Objective Measures of Presenter Performance ................................................................. 85

6.3.6 Qualitative Findings ........................................................................................................ 87

6.4 CONCLUSION ....................................................................................................................... 89

ROBOCOP SYSTEM ............................................................................................................... 90

7.1 UNDERSTANDING REHEARSAL COACHING PRACTICE .............................................. 91

7.1.1 Participants ....................................................................................................................... 91

7.1.2 Procedure ......................................................................................................................... 91

7.1.3 Findings ............................................................................................................................ 92

7.2 DESIGN OF ROBOCOP ...................................................................................................... 94

7.2.1 Presentation Preparation with RoboCOP ......................................................................... 95

7.2.2 The Robotic Coach ........................................................................................................... 96

7.2.3 Presentation Quality Analysis ......................................................................................... 97

7.2.4 Identifying Performance Trends ....................................................................................... 100

7.2.5 Feedback Generation ....................................................................................................... 101

7.3 COMPARISON OF FEEDBACK MODALITIES .................................................................. 101

7.4 EVALUATION OF ROBOCOP ............................................................................................ 102

7.4.1 Procedure ......................................................................................................................... 103

7.4.2 Presenter Participants ....................................................................................................... 105

7.4.3 Presenter Measures .......................................................................................................... 105

7.4.4 Presenter Quantitative Results ......................................................................................... 105
SLIDES USED IN INTELLIPROMPTER STUDY ............................................................... 147
D.1 ITALIAN ART ........................................................................................................... 147
D.2 ITALY TOURIST ATTRACTIONS ............................................................................ 150
D.3 ITALIAN CUISINE .................................................................................................. 153

STUDY PROTOCOLS FOR QUESTER STUDY ............................................................... 156
E.1 PROTOCOL FOR FORMATIVE STUDY .................................................................. 156
E.2 INTERVIEW GUIDE FOR FORMATIVE STUDY .................................................... 157
E.3 PROTOCOL FOR QUESTION GATHERING STUDY ............................................. 159
E.4 PROTOCOL FOR EVALUATION STUDY ............................................................. 161
E.5 INTERVIEW GUIDE FOR EVALUATION STUDY ................................................ 165

QUESTIONNAIRES FOR QUESTER STUDY ................................................................ 167
F.1 QUESTION ANSWERING (QA) RATING QUESTIONNAIRE .................................. 167
F.2 PRESENTATION SELF-RATING QUESTIONNAIRE ............................................. 168
F.3 RANKING OF THE TELEPROMPTER SYSTEMS ................................................... 169

SLIDES USED IN QUESTER STUDY .......................................................................... 170
G.1 NUTRITION AND DIET .......................................................................................... 170
G.2 QUESTIONS REGARDING NUTRITION AND DIET SLIDES .............................. 192
G3. PHYSICAL EXERCISE .............................................................................................. 194
G4. QUESTIONS REGARDING PHYSICAL EXERCISE SLIDES ................................ 216

BIBLIOGRAPHY .......................................................................................................... 218
List of Tables

TABLE 1. STATISTICS GATHERED FROM THE NOTES CORPUS ................................................................. 27

TABLE 2. KEYWORDS AND WEIGHTS FROM SEGMENT 3 OF SAMPLE NOTES .............................................. 44

TABLE 3. TRACKING METHODS EVALUATED IN STUDY ..................................................................................... 48

TABLE 4. EVALUATION MEASURES USING THRESHOLDS OPTIMIZED FOR BEST F-SCORE ................................. 50

TABLE 5. STATISTICAL ANALYSIS OF CONFERENCE PRESENTATIONS AND LECTURES ................................. 72

TABLE 6. PRESENTERS’ SELF-PERCEIVED RATINGS OF THEIR QA PERFORMANCES (MEAN (SD) AND P-VALUE OF WILCOXON TESTS)  ........................................................................................................... 85

TABLE 7. OBJECTIVE MEASURES OF PRESENTERS’ SYSTEM USAGE AND PERFORMANCE DURING QA SESSIONS (MEAN (SD) AND P-VALUE OF WILCOXON SIGNED RANK TESTS) ................................................................. 87

TABLE 8. COMMON FEEDBACK CATEGORIES AND THEIR FREQUENCIES ............................................................ 93

TABLE 9. RANGES AND THRESHOLDS FOR SPEECH QUALITY METRICS .......................................................... 100

TABLE 10. AVERAGE RATINGS OF THE ROBOTIC COACH .................................................................................... 106

TABLE 11. PRIMARY OUTCOMES OF THE STUDIES ............................................................................................ 118
List of Figures

FIGURE 1. THE AUTHORING ENVIRONMENT AND PRESENTER VIEW IN MICROSOFT POWERPOINT ........................................ 2

FIGURE 2. IN THIS DISSERTATION, I EXPLORE THE APPLICATIONS OF PRESENTATION TRACKING TO SUPPORT PRESENTATION

REHEARSAL (CHAPTER 7 - THE ROBOCOP SYSTEM), PRESENTATION DELIVERY (CHAPTER 5 - THE INTELLIPROMPTER

SYSTEM), AND THE QUESTION-ANSWERING SESSION AFTER PRESENTATION (CHAPTER 6 - THE QUESTER SYSTEM) .......... 6

FIGURE 3. VISUAL MAPPING OF THE SPACE OF PRESENTATION TYPES ........................................................................... 10

FIGURE 4. PITCHPERFECT SYSTEM, TRINH ET AL. [112] ................................................................................................... 11

FIGURE 5. ALIGNING UTTERANCE CLUSTERS WITH SLIDE SUBSECTIONS, LU ET AL. [63] ............................................. 15

FIGURE 6. SAMPLE RECOGNITION LATTICE AND CORRESPONDING MULTIPLE ALIGNMENT REPRESENTED AS CONFUSION NETWORK.

MANGU ET AL. [67] .................................................................................................................................................. 19

FIGURE 7. VECTOR DIFFERENCES OF GLOVE [84] VECTOR REPRESENTATIONS OF WORD PAIRS WHICH ARE DISTINGUISHED WITH

THE SAME UNDERLYING CONCEPTS IS ROUGHLY EQUAL. (FIGURE RETRIEVED FROM

HTTPS://NLP.STANFORD.EDU/PROJECTS/GLOVE/) ................................................................................................. 22

FIGURE 8. DYNAMICDUO NOTE AUTHORING ENVIRONMENT: A) NOTES FOR AGENT AND HUMAN PRESENTER. B) A NOTE

SUBSECTION ASSIGNED TO THE AGENT. C) PREVIEW OF THE AGENT. TRINH ET AL. [111] .............................................. 36

FIGURE 9. A PARTICIPANT CO-PRESENTING WITH THE VIRTUAL AGENT ........................................................................ 37

FIGURE 10. OVERALL PRESENTATION TRACKING FRAMEWORK ...................................................................................... 40

FIGURE 11. SAMPLE SLIDE NOTES WITH KEYWORD CANDIDATES IN BOLD. .................................................................... 44

FIGURE 12. A SAMPLE KEYWORD SPOTTING SCENARIO. ................................................................................................. 45

FIGURE 13. PRECISION-RECALL CURVES FOR DIFFERENT TRACKING METHODS .......................................................... 49

FIGURE 14. AUTHORING ENVIRONMENT (LEFT) AND PRESENTER VIEW (RIGHT) ............................................................. 53

FIGURE 15. TOPIC ORDER AND SCROLLING PROCESS EXAMPLE. RED RECTANGLE SHOWS THE VISIBLE SECTION ON NOTES. A) THE

FIRST TOPIC IS HIGHLIGHTED AND DISPLAYED AT THE TOP. B) TOPIC 1 IS COVERED; A TICK MARK APPEARS NEXT TO TOPIC 1,
AND TOPIC 2 IS HIGHLIGHTED. C) NOTES SCROLL UP TO PLACE TOPIC 2 AT THE TOP. D) TOPIC 3 IS COVERED, TOPIC 2 IS
MOVED TO THE BOTTOM. E) NOTES SCROLL UP TO PLACE TOPIC 4 AT THE TOP.................................................................................................................. 55

FIGURE 16. A PARTICIPANT PRESENTING WITH GOOGLE GLASS................................................................................................................................. 57

FIGURE 17. ABSOLUTE RATINGS OF THREE NOTE DISPLAY SYSTEMS (* INDICATES STATISTICAL SIGNIFICANCES IN PAIRWISE
COMPARISONS) .............................................................................................................................................................................................. 58

FIGURE 18. NOTE AUTHORING ENVIRONMENT (LEFT) AND PRESENTER VIEW (RIGHT): (A) NOTES PANE WITH AN EXTRA NOTE
TOPIC IN RED, (B) SLIDE NAVIGATION PANE SHOWING THE MOST RELATED SLIDES, C) GREEN BARS SHOWING THE DEGREE
OF SLIDE RELATEDNESS TO SPEECH INPUT ................................................................................................................................................. 73

FIGURE 19. (LEFT) SLIDE PREVIEW DISPLAYED WHEN THE PRESENTER HOVERS THE MOUSE OVER A SLIDE THUMBNAIL; (RIGHT)
NOTE PANE IS SCROLLED TO DISPLAY THE MOST RELEVANT NOTE TOPIC AT THE TOP WITH HIGHLIGHTED KEYWORDS ...... 74

FIGURE 20. SAMPLE AUTHORED SLIDE AND THEIR RELATED QUESTIONS.......................................................................................................................... 77

FIGURE 22. PERFORMANCE OF THE DECAY RECENCY MODEL WITH DIFFERENT BASE VALUES......................................................... 80

FIGURE 21. COMPARISON OF DIFFERENT RECENCY MODELS .............................................................................................................................. 80

FIGURE 23. PRESENTERS’ ABSOLUTE USABILITY RATINGS OF TWO QA SUPPORT SYSTEMS (* INDICATES STATISTICAL
SIGNIFICANCES) .............................................................................................................................................................................................. 83

FIGURE 24. PRESENTER REHEARSING HER TALK IN FRONT OF THE ROBOTIC COACH ............................................................................................ 90

FIGURE 25. ROBOCOP MICROSOFT POWERPOINT ADD-IN WITH (A) NOTE AUTHORING PANE AND (B) PRESENTER VIEW ...... 94

FIGURE 26. EXAMPLE OF GRAPHICAL FEEDBACK DISPLAYED AT THE END OF EACH SLIDE ................................................................. 102

FIGURE 27. REHEARSAL SETUP FOR (A) ROBOCOP AND (B) VOICE ONLY & GRAPHIC CONDITIONS IN THE FEEDBACK MODALITY
STUDY........................................................................................................................................................................................................ 103

FIGURE 28. ABSOLUTE RATINGS OF PRESENTATION QUALITY FOR THE ROBOT VS. CONTROL CONDITIONS (* INDICATES
SIGNIFICANT DIFFERENCES) ............................................................................................................................................................................. 107

FIGURE 29. AUDIENCE PERCEPTION OF PRESENTERS FOR THE ROBOT VS. CONTROL CONDITIONS (* INDICATES SIGNIFICANT
DIFFERENCES)................................................................................................................................................................................................... 107

FIGURE 30. RELATIVE RATINGS OF PRESENTATION QUALITY FOR THE ROBOT VS. CONTROL CONDITIONS (POSITIVE VALUES
INDICATE THE ROBOT CONDITION IS BETTER, * INDICATES SIGNIFICANT DIFFERENCES) ................................................................. 109

FIGURE 31. SUMMARY OF THE CONTRIBUTIONS OF THIS DISSERTATION ..................................................................................................... 117
CHAPTER 1

Introduction

With millions of oral presentations being made everyday, public speaking is a critical component of our academic, scientific and professional life, serving as a powerful communication channel to inform, inspire and persuade audiences. Given the significant role of public speaking skills in our academic and career success [23], it is unsurprising that it has long become the main subject of basic communication courses in the majority of universities in the United States [75]. Despite the importance and pervasiveness of oral presentations, the average quality of professional presentations is still low, as evidenced in an extensive survey of 2,501 professionals [41]. In this survey, respondents gave an average grade of “C-” (2.9 on a 1-to-5 scale) for presentations they had attended. There are several verbal and non-verbal factors that affect the quality of presentations, including content mastery, proper organization and transitions, the use of vocal variety to engage the audience, the expression of confidence through postures, gestures and eye contact, and the effective use of visual aids [99]. Reasons for poor presentation quality include: insufficient preparation and rehearsal time, lack of skill training, and public speaking anxiety (affecting at least 35% of the population) [12, 94]. Specifically, presentation rehearsal has been identified as a significant predictor of presentation quality, as it enables the speaker to more fully develop their perspective-taking and audience-analysis skills than in solo rehearsal [70].
1.1 Current Slideware Technology

Since the first release of Microsoft PowerPoint\(^1\) in 1987, several slideware products (including Apple Keynote\(^2\)), have been publicized and widely used to support the authoring, rehearsal, and delivery of oral presentations. These systems are designed based on the slide metaphor [31], in which presentation media are broken down into a series of discrete units called slides, each of which may contain a variety of visual elements, such as text, pictures and charts (Figure 1a). During presentation delivery, these slides are displayed to the audience in a fixed, linear order to aid the presenter’s speech. In addition to the slide visual design, most slideware systems also enable presenters to plan their verbal content through the authoring of speaking notes for each slide. During the delivery, these notes can be displayed in a private ‘Presenter View’ screen, which is accessible only to the presenters to remind them of what to say for each slide (Figure 1b).

![Figure 1. The Authoring Environment and Presenter View in Microsoft PowerPoint](image)

1.2 Problems of Current Slideware

Xingeng et al. [119] identifies some of the disadvantages of using slideware technology, including: note reliance (in the extreme case leading to presenters reading their speech verbatim) and increased

\(^1\) https://products.office.com/en-us/powerpoint

\(^2\) https://www.apple.com/keynote/
focus on the screen (both of which reduce the presenter’s eye contact), too much information on the slides (which makes the presentation more difficult to follow), and fixed ordering of slides which may restrict improvisation. Tufte [113] argues that PowerPoint encourages presenters to compress ideas into bullet points, while disregarding the narrative between the points, and arrange them into avoidably deep hierarchies using poorly designed templates. Fear of forgetting what to say next might encourage dense content on slides, which can lead to presenters reading from the slides rather than making connection with the audience [28]. Reynolds [91] discourages reading the text during the presentation and instead recommends using short lines as cues to key points, which can be glanced at during the presentation.

Current slideware technologies (such as PowerPoint) provide very rudimentary note displays in the “Presenter View” mode. However, these displays are not dynamically updated to show the note section most relevant to the presenter’s speech, requiring presenters to either manually scroll through the notes or condense all of their notes into a few keywords or phrases. Even then, they may have to scan their notes periodically to find the next relevant cue. “Presenter View” notes are also typically displayed on the presenter’s podium laptop, requiring them to break eye contact with their audience when referencing their notes.

Another issue with the current slideware systems is that they are primarily designed to support presenters in delivering standard, scripted, linear, and rehearsed presentations in a polished, but fixed and invariant manner [28, 113, 119]. Unfortunately, these tools do not provide good support for dynamic, free-form presentations, which can encourage discussions, creative thinking, and communication of new ideas [59]. Tufte [113] criticizes the linear structure of PowerPoint presentations, arguing that it is not “audience-oriented” and can reduce the overall quality of the presentations. The “sequentiality” of slides makes it difficult to navigate to related slides in arbitrary orders, and thus may discourage presenters from dynamically adapting their presentation.
in order to respond to audience questions and evolving audience needs [124]. At best, current slideware systems only provide simple indexing mechanisms to let speakers hunt through their slides for material to support their dynamically evolving speech, and speakers must perform this frantic search while the audience is watching and waiting.

Finally, commercial slideware systems only provide limited features to aid presentation rehearsal, including display of presentation timing and audio recording of the presenter’s speech. There is a lack of coaching support to provide feedback on various verbal and non-verbal behaviors that could impact the quality of oral presentations.

1.3 Presentation Tracking and Potential Applications

Utilizing the content of the presenter’s speech in applications related to oral presentations is a complex task, due to the dependency on accuracy of automatic speech recognition and the fact that speaker rarely follows presentation notes exactly. Previous studies either provide very limited support for deviations from presentation script [78] or only perform off-line alignment of speech and presentation content [61, 88]. Real-time tracking of presentation content is a novel approach for providing content-based assistance during presentations. Given the content of presentation slides and speaking notes, the goal of a presentation tracking system is to track the speaker’s content coverage relative to their supporting media in real-time while allowing for flexibility in speakers’ wording usage and order of presentation content. Applications equipped with this feature can provide better content-based assistance during presentation rehearsal and delivery.

For instance, the tracking information can be used to provide intelligent teleprompters [7]. Effective teleprompting can serve to reduce speaking anxiety by alleviating speakers of the fear of forgetting important points, or even the overall structure of their talk. Teleprompting also allows presenters to minimize the text they put on their slides as memory aides for themselves, allowing the audience
to focus on the speaker instead of dividing attention between the speaker and reading slide text. This can lead to better usage of slides as visual aids for the audience and help the presenter to avoid ineffective slide authoring practices previously criticized by Tufte [113]. However, most existing teleprompter systems require either a human operator or presenters themselves to manually scroll through the note text, or simply advance the notes based on a fixed scrolling speed. In contrast, an intelligent teleprompter system could use presentation tracking to automatically navigate to the relevant note segment and highlight key phrases to remind the presenter what to say next.

Another potential application of presentation tracking is in facilitating dynamic and nonlinear access to relevant presentation slides and supporting media based on speech input. This can provide good support for speakers who want to deliver more extemporaneous talks, in which they dynamically adapt their presentation to input or questions from the audience, evolving audience needs, or other contextual factors such as varying or indeterminate presentation time, real-time information, or more improvisational or experimental formats. Perhaps the most common situations in which speakers must provide such dynamic presentations are in Question and Answer sessions. Speakers' remarks are entirely driven by the questions from the audience, which arrive in an unscripted order, and the questions themselves are often entirely unanticipated. A significant portion of these Question & Answer sessions is often taken up by the speaker searching through their slides for material to support their responses. Several solutions have been developed to support non-linear, dynamic presentations [29, 76, 104]. However, they require authoring of pre-planned paths and multiple branches in presentation scripts. The use of presentation tracking could reduce the need for prior preparation, by dynamically offering branching options in real time based on tracking of the presenter’s spoken utterances.

Presentation tracking can also be employed to provide greater support for presenters during presentation rehearsal. There have been a number of rehearsal support systems that provide
feedback on different aspects of presentation performance, such as speaking rate, vocal variety and eye contact [8, 16, 98]. Presentation tracking can supplement rehearsal support systems by providing content-based feedback.

![Diagram](image)

**Figure 2.** In this dissertation, I explore the applications of presentation tracking to support presentation rehearsal (Chapter 7 - the RoboCOP system), presentation delivery (Chapter 5 - the IntelliPrompter system), and the question-answering session after presentation (Chapter 6 - the Quester system)

### 1.4 Contributions

In this dissertation, I present a framework for real-time oral presentation tracking through semantic matching. The thesis of this dissertation is that utilizing presentation tracking in presentation assistance applications can improve the experience and performance of presenters during rehearsal and delivery of presentations and QA sessions (Figure 2).

Specifically, I tackle the following research questions:

1. How can speech recognition and semantic text retrieval techniques that have been successfully applied in similar applications be adopted to develop a real-time oral presentation tracking framework?

2. To what extent does an automatic teleprompter that utilizes presentation tracking improve the presenters’ experience?
3. When would presentation tracking be effective in supporting dynamic presentations?

4. How effective is presentation tracking in providing content-based feedback during presentation rehearsal?

Through several exploratory and evaluation studies, I answer these questions and provide the following contributions:

- The results of exploratory studies of experienced presenters to determine current practices in preparing and delivery of oral presentations;
- Design of a real-time oral presentation tracking framework, based on a requirement analysis study and evaluation of the performance of different implementations of the framework;
- Evidence for the effectiveness of presentation tracking in applications regarding presentation rehearsal and delivery based on results of user studies with a teleprompter system, a question-answering support system, and a public speaking coach system.

1.5 Dissertation Structure

This dissertation is organized as follows. In Chapter 2, I review related work on presentation assistance systems and examine common techniques in speech recognition and semantic text retrieval.

In Chapter 3, I analyze interviews with expert presenters regarding common practices and requirements for non-linear presentations and usage of presentation slide notes.

In Chapter 4, I present the requirement analysis, design, implementation and evaluation of a real-time oral presentation tracking framework (published in [3]).

In Chapter 5, I describe an intelligent teleprompter system which incorporates presentation tracking and present the results of user evaluation studies (published in [4]).
In Chapter 6, I present the design and user evaluation of a speech based question-answering support system for oral presentations (published in [5]).

In Chapter 7, I examine the effectiveness of tracking for rehearsal applications. I describe a robotic coach for oral presentation which uses presentation tracking to provide feedback on the presentation content coverage, and present the results of user evaluations of this system (published in [110]).

In Chapter 8, I summarize this dissertation work and explore the directions for future research.
CHAPTER 2

Background and Related Work

In this chapter, I first define the space of presentation types that are the target of this dissertation, before presenting a brief survey of currently available presentation assistance systems and the limited work on presentation tracking. Finally, I explore previous work on two main challenges in presentation tracking: automatic speech recognition and semantic text retrieval.

2.1 Presentation Typology

Presentation activities can be categorized based on different attributes, such as audience size, power relationship between presenter and audience, presentation goals, and forms of visual support. Since the focus of this dissertation is on matching speech input with presentation content, I am mostly interested in attributes related to the order and wordings of presentations. I define a visual mapping of the space of presentation types with two dimensions: Linearity (presentation order matters) and Scriptedness (precise wording matters). Figure 3 displays this visual mapping, along with some examples of different presentation types in this space: extemporaneous talks (low linearity, low scriptedness), classroom lectures (high linearity, low scriptedness), formal interviews (low linearity, high scriptedness), and formal speech (high linearity, high scriptedness).

Presentation tracking requires some degree of scriptedness, since it utilizes pre-planned presentation content. Presentations with high scriptedness and linearity can be readily supported by existing, less complex systems such as traditional teleprompters, which do not need to be robust against deviations from the exact wordings and order of presentation content. Therefore, in this
dissertation, I target an area that has not been adequately addressed in previous work – that is presentations with intermediate degree of linearity and scriptedness (Figure 3).

![Figure 3. Visual mapping of the space of presentation types](image)

In the following sections, I will refer to this mapping to describe the presentation activities supported by previous presentation support systems and discuss their limitations.

### 2.2 Presentation Assistance Technologies

#### 2.2.1 Presentation Authoring Support

Several presentation tools have been developed to provide support for the planning and authoring of slide content and speaking notes. Liu et al. [61] developed the SidePoint add-in for PowerPoint, a peripheral panel that automatically suggests relevant knowledge items based on the textual content of the current slide. The system enables the presenter to import desired items (e.g. related facts or descriptions) into the note section of the current slide. Pschetz et al. [87] designed TurningPoint, a presentation planning environment that enables the presenter to brainstorm both visual and verbal content using sticky-like notes and sequence notes into a narrative structure using
narrative templates. Once a narrative is created, the system automatically generates a corresponding slide deck by adding the planned visual and verbal content into the PowerPoint slide and note sections respectively. Although not directly related to presentation tracking, these systems present novel approaches to the creation of slide content and supporting notes, which could inform the design of a presentation tracking framework and its potential applications.

### 2.2.2 Presentation Rehearsal Support

Recent research has addressed the need for more effective approaches to presentation rehearsal. Trinh et al. [112] developed the PitchPerfect system, which provides an integrated rehearsal environment with a range of targeted rehearsal tools for structured presentation preparation (Figure 4). To support the planning of verbal delivery, the system incorporates an extended authoring tool that enables the presenter to annotate each visual element on a slide with speaking notes. This sub-slide note division approach could potentially be adopted to facilitate tracking of presentation note coverage at a more fine-grained level than is supported by slide-level notes.

![Figure 4. PitchPerfect System, Trinh et al. [112]](image-url)
Several public speaking training platforms have also been developed that provide feedback on different aspects of presentation delivery, from speech quality to speaker body language. Kurihara et al. [52] developed the Presentation Sensei system, which provides graph-based visual feedback on the presenter’s speaking rate, eye contact, filler rate and timing. Tanveer et al. [108] designed the AutoManner system, which offers visual feedback on the speaker’s body movements. Lui et al. [64] developed a mobile application that displays feedback on body motion, voice intensity and timing. Schneider et al. [98] developed the Presentation Trainer system, which generates both visual and haptic feedback on the speaker’s voice intensity, use of pauses and fillers, body posture and hand gestures. Similarly, the AwareMe system [16] measures voice pitch, filler words, and speaking rate during presentation rehearsal and provides visual and haptic feedback through a wristband device.

Previous studies have also explored the use of virtual agents to facilitate practice of communication skills. Batrinca et al. [8] developed Cicero, a virtual audience platform for public speaking training. The virtual audience can display indirect, nonverbal feedback to signal increased attention, cues of rapport, lack of interest, or disagreement in response to sensed speaker behaviors. Although not designed for public speaking, Hoque et al. [44] developed a related system to provide automated job interview training. The MACH system uses a highly realistic animated virtual job interview coach to offer real-time visual feedback on various verbal and nonverbal behaviors of human interviewees, including speech, prosody and facial expressions.

Evaluation results of existing public speaking training systems often demonstrate the positive effects of automated feedback on increasing user engagement and improving learners’ experiences. However, most of these systems have only focused on generating feedback on speech quality and speakers’ nonverbal behaviors. There have been no prior studies that explore the use of presentation tracking to provide content-based feedback during rehearsal.


2.2.3 Presentation Delivery Support

Several speech assistance studies aim at providing support during presentation delivery. Saket et al. [95] designed a mobile application for timing support during the presentation. Tam et al. [106] developed a wireless wrist-worn system which provides haptic feedback for time-management.

Head mounted displays such as Google Glass have been used for providing feedback during presentation. Rhema, for example, [20] is a presentation assistance interface on Google Glass that provides real-time feedback on speech volume and speaking rate. In a comparative study, speakers preferred sparse feedback compared to continuous feedback and to no feedback conditions. Logue [3] is another system that provides real-time feedback on the speaker’s energy, openness, and speaking rate on a head-mounted device. Speakers found the system to be helpful and not distracting, and observers found the users’ postures more open. Head mounted devices have also been used to display cue cards to actors for cold reading [23] and to support MC performers [14]. Findings from these studies suggest that head-mounted devices could potentially be a viable modality for displaying real-time presentation tracking information to the presenter during delivery.

Previous studies have also explored the use of virtual characters as presenters. Noma et al. [80] developed a virtual presenter that is capable of presenting authored text and non-verbal behaviors. DynamicDuo [111] provides a virtual agent as a co-presenter. The agent can deliver pre-assigned parts of the presentation at the user’s request. A user study showed that using this system resulted in significant decrease in users’ anxiety, increase in their confidence, and higher audience ratings, compared to solo presentations. DynamicDuo, however, only supports fully pre-scripted and linear co-presentations (high scriptedness and high linearity), requiring the human and agent presenters to strictly follow their planned sequence of turn-takings during delivery. The integration of presentation tracking capability could allow for a truly dynamic human-agent collaboration, in which the agent co-presenter could monitor which segments of slide notes that have been deliver
by the human presenter in real-time and proactively update the initial co-presentation plan accordingly.

2.2.4 Non-Linear Presentation Support

Previous research has developed novel presentation systems to support a more dynamic approach to presenting. One of the challenges in designing these systems is to avoid distraction and excessive cognitive load that dynamic interfaces may cause [36]. Moskovich et al. [7] developed an interface for authoring and delivering customizable presentations using a hierarchical presentation structure and predefined paths. During delivery, the presenter can change the presentation path based on audience feedback or time constraints. The NextSlidePlease system [10] provides a graph-based authoring system and a presentation path suggestion algorithm to support interactive presentations. It helps users prioritize presentation content based on remaining time during presentation delivery. The HyperSlides system [5] supports authoring of hyperlinked, non-linear presentation slides. Finally, the Palette [79] and PaperPoint [101] systems use paper-based slide cards to provide random access to single digital slides, thereby enabling the presenter to easily deliver slides in a non-linear and flexible fashion.

One of the main challenges in providing automated support for non-linear presentations is capturing speakers’ dynamic and evolving communicative intent during delivery. Inferring intent from the tracking of presenters’ speech content is an ideal approach, since it does not involve additional effort from the speaker to either author multiple possible branches in presentation scripts or prepare special physical slide cards, as required in previous systems.
2.2.5 Presentation Tracking

Limited work has been done leveraging the content of the presenter’s speech. One possible reason is the dependency of such systems on the accuracy of speech transcriptions, but with current advances in automatic speech recognition this area has gained more attention. Presentation tracking using speech content can facilitate automatic slide transitions, or can remind the user of the presentation flow, in order to reduce note reliance. Rogina et al. [92] developed a lecture tracker which can be used to switch slides and display the documents related to speech. They used dynamic time warping to match slides with the ASR output hypothesis. They achieved about 30% improvement in tracking error rate compared to a baseline method that assigned each slide a time slot of the same size. Okada et al. [81] computed the minimum distance between ASR hypothesis and speech script in order to track the current state of speech in real-time. The tracking system was designed for supporting master of ceremony (MC) performances and assumed that the speech performance would be very close to the planned speaking notes (high scriptedness).

![Figure 5. Aligning utterance clusters with slide subsections, Lu et al. [63]](image)

Some of the studies related to audio lecture indexing and retrieval use techniques which can be used for tracking. Lu et al. [63] used entropy-based word filtering, reliability-propagated word-based matching, and structured support vector machines to align utterance clusters with slide subsections (Figure 5). To my knowledge, this is the only work to date on alignment within slides;
however, it assumes in-order presentation (high linearity), and is an off-line system. Yamamoto et al. [122] compared the results of ASR with the textbook used in the lecture to segment lecture recordings into topics. To do so, they calculated the similarity between word vectors from speech transcript and textbook. Yang et al. [123] extracted the keywords from video lecture by using the ASR output and text information extracted from videos. Cao et al. [19] associated the text extracted from PowerPoint slides with the lecture videos to support video-based question answering on lecture videos.

To summarize, previous work on speech-based tracking, archiving, indexing, and retrieval of lectures and speech scripts has provided helpful insights into the use of ASR and text retrieval techniques to support the development of a presentation tracking framework. However, existing systems rely on either linearity of slides and topics within a slide, or exact word matching between spoken utterances and pre-planned scripts (high scriptedness). Thus, they only provide limited opportunity for the presenter to improvise during delivery. In this thesis, I aim to address these limitations of the existing systems by exploring novel real-time presentation tracking approaches that can allow for flexibility in the slide and topic orders, while accommodating a much greater degree of improvisation in the presenter’s speech.

In the following sections, I present a more detailed review of speech recognition and text retrieval methods, which have previously been identified as two main challenges in presentation tracking.

### 2.3 Speech Recognition

Automatic speech recognition (ASR) is a critical part of speech based interfaces which is responsible for transcribing the users’ speech input. Speech recognition has improved significantly from single-speaker digit recognition systems in 1952 [50] to speaker-independent continuous speech recognition systems based on deep neural networks [43]. Currently, several open source
ASR engines such as Pocketsphinx [46], Kaldi [86], and HTK [118] are available, but accurate speech recognition requires high processing power which cloud based services such as IBM Watson¹, and the Google cloud platform² provide. Since ASR is the first operation performed in speech based pipelines, errors in speech recognition can often result in major reductions in accuracy of the overall system. Recent ASR systems have decreased the word error rate to around 5% [97, 120], which has resulted in increased usage of speech input in wide range of applications. Speech recognition has been used in several applications related to audio lectures including transcription, indexing, and retrieval [37, 90, 109].

2.3.1 Acoustic and Language Models

Almost all of the speech recognition systems use acoustic and language models and have a vocabulary which contains the words that can be recognized by them [88].

*Acoustic Model:*

Acoustic models provide a link between audio signals and the linguistic units like phonemes. They are generated from databases of speech audio samples and their transcriptions, such as TIMIT [33] and SWITCHBOARD [38]. Speech corpora generally have low diversity of speakers, therefore acoustic models generated from them might be inaccurate for transcribing speech input from non-native speakers, speakers with accents, speakers affected with speech impairments [10], or others underrepresented in the corpora, such as older adults and children. Also, recording factors such as noise and other audio distortions can result in lower ASR performance [57].

¹ https://www.ibm.com/watson/services/speech-to-text/

² https://cloud.google.com/speech/
**Language Model:**

Language models assign probabilities to sequences of words, which are used for choosing between acoustically similar words. Factors such as disfluencies, short duration, and being at the start of a turn can result in misrecognized words [39]. Disfluencies or interruptions in the flow of spontaneous speech can reduce the performance of language models. They include filled pauses (tokens like “uh”, “er”, “um” that are used to fill time), repetitions, and false starts [18]. Another source of error comes from occurrence of uncommon words which are not included in the ASR system vocabulary, called out-of-vocabulary (OOV) words [9]. Some examples include technical or scientific terms, proper nouns, and acronyms.

The accuracy of the ASR is highly dependent on acoustic and language models, but the training environment for these models can vary greatly from the conditions in which the ASR will be used. In such cases, methods such as acoustic model adaptation [117] and language model adaptation [22] can improve the ASR performance. Preprocessing the ASR output to detect disfluencies before passing to the language model can also reduce the error rate [126].

Previous studies have shown that the accuracy of the ASR can be improved for lecture transcription by retrieving text related to the presentation from the web or other supplementary material, and adapting the vocabulary and language model used for speech recognition. Park et al. [82] used a combination of spontaneous speech resources and textbooks as the language model. Munteanu et al. [77] used the contents of all the slides in the lecture to compile a corpus which eliminated the need for two different general and topic specific language models. Maergner et al. [66] used feature-based ranking for vocabulary selection. They generated a vocabulary using a collection of documents that were similar to lecture slides, and then ranked the resulting vocabulary based on a combination of word features.
2.3.2 Confusion Networks

In some cases, such as using cloud-based ASR systems, we have limited control over the acoustic and language models used for ASR, and we cannot use the approaches mentioned above to improve the accuracy of the ASR. Alternatively, to deal with imperfect ASR, we can reduce the vulnerability of the overall system to ASR errors. To do so, instead of using only the best hypothesis from the ASR system, multiple ambiguous hypotheses are processed. These hypotheses, in the form of an ASR output graph, are called a confusion network or lattice [67] (Figure 6).

![Image of a sample recognition lattice and corresponding multiple alignment represented as confusion network. Mangu et al. [67]](image)

Fuji et al. [35] proposed using confusion networks to improve the robustness against recognition errors. They achieved 8.9% improvement in word error rate (WER) compared to using 1-best results. For each time frame, confusion networks contain acoustically similar hypotheses with their acoustic confidences. This rich information has been used in many speech-related applications such as speech translation [11], semantic parsing [114] and spoken language understanding [71]. In an application similar to presentation tracking, Hori et al. [45] used confusion networks for spoken
utterance retrieval from a lecture corpus. They performed keyword matching on out of vocabulary (OOV) words by combining phone and word confusion networks.

In this work, I aim to evaluate the performance of different ASR systems and methods to improve the robustness of ASR, including the use of confusion networks, in order to support accurate real-time presentation tracking.

### 2.4 Semantic Text Retrieval

To measure the coverage of slide content for presentation tracking, a system should associate ASR output with related segments of slide content without reliance on order or exact wordings of presentation content. Text retrieval studies focus on the more generalized problem of matching queries against a set of text documents. A common approach in text retrieval is to extract features from text, assign weights to them based on their importance in the text, and use these weights in the decision-making process. Proper weighting methods have been shown to be more important than the feature selection process [17].

#### 2.4.1 Term Weighting

Term frequency-inverse document frequency (tf.idf) is a common term weighting method initially proposed by Jones [103], in which terms are weighted based on their frequency and their specificity to a document in a collection of documents. Salton et al. [96] compared different automatic term weighting methods. In their work, the query and document sets were represented by vectors containing all terms and their assigned weights. The similarity of query to documents was measured by vector similarity functions such as the cosine vector similarity formula. Different term frequencies, collection frequencies, and normalization factors were used for assigning weights to words. A normalization factor was used to remove the advantage of long documents which have higher term frequencies and more words.
Term weighting is also the subject of studies on keyword extraction and text summarization. Fujii et al. [34] automatically summarized lecture speech by extracting cue phrases using Conditional Random Fields (CRF). Cue phrases were defined as word sequences that appear in important sentences more than non-important sentences. A CRF was trained using a dataset in which Important/non-important sentences were manually labeled and cue phrases were extracted. Kawahara et al. [78] extracted characteristic keywords of the lecture using tf-idf weighting and then used the extracted keywords as one of the measures for indexing key sentences in lecture archives. Yang et al. [123] only considered nouns and numbers as keyword candidates and then took the top N words ranked by the word frequency. They weighted the words using an extended version of tf.idf that also considers the location information of keywords.

### 2.4.2 Semantic Similarity

Semantic text retrieval considers the meaning of text, which can be lost in general statistical methods [51]. Meaning of words is especially important in processing the presentation speech transcripts, since speakers often utter words that are semantically related to the text keywords without speaking any of the exact keywords [54]. In [62], the authors argued that good keywords should be semantically relevant to the document theme and also provide a good coverage of the concepts. They clustered terms based on their semantic relatedness and extracted key phrases from exemplar terms in these clusters.

Metrics for measuring the semantic similarity of words can be put in two categories [72]: corpus based measures, which use the information gathered from large corpora, e.g., word co-occurrence [115], and knowledge based measures, which use information from semantic networks such as WordNet [74]. Islam et al. [47] combined corpus based co-occurrence metric with string similarity metrics to measure the semantic similarity. Mihalcea et al. [72] defined the semantic similarity between two text segments by combining the semantic similarities of
each text segment with the other one. They identified and measured the similarity of the most similar word in each segment to each word in the other segment. Then, they used word specificity to assign weights to these similarity values and normalized them based on the length of sentence.

Mikolov et al. [73] used continuous vector representations of words, computed from a very large dataset using neural networks, to measure word semantic similarity. The vectors representing words that are semantically close in the dataset are located close to each other in the vector space. Pennington et al. [84] proposed Global Vectors for word representation (GloVe) and

![Figure 7. Vector differences of GloVe [84] vector representations of word pairs which are distinguished with the same underlying concepts is roughly equal. (Figure retrieved from https://nlp.stanford.edu/projects/glove/)](https://nlp.stanford.edu/projects/glove/)
claimed that their methods outperform other word embedding methods in text similarity tasks. GloVe vectors were trained using non-zero elements of a word-word co-occurrence matrix gathered from a large corpus (Figure 7). Yih et al. [125] measured the semantic similarity of entities and relations extracted from questions with the content of the knowledge base to perform open domain question answering. Semantic parsing was performed using a neural network-based semantic model.

In this dissertation, I aim to investigate the effectiveness of different weighting techniques, including both statistical-based term weighting and semantic similarity methods, with the goal to develop a robust presentation tracking framework that allows for great flexibility in speakers’ wording usage during delivery.

2.5 Conclusion

In this chapter, I examined previous work related to presentation tracking. I reviewed a number of presentations assistance systems that provide support for speakers during the authoring, rehearsal, and delivery of presentations. These systems proposed novel approaches to generating and structuring presentation content and speaking notes, which could inform the design of a presentation tracking framework. This analysis also demonstrated the lack of studies on real-time presentation tracking, and highlighted several potential applications of tracking to provide feedback on the content of the presenter’s speech and offer greater support for presentations with intermediate amount of scriptedness and linearity. Finally, I explored related studies on automatic speech recognition and text retrieval, which are key to the development of a presentation tracking framework. I identified methods that can be adopted to improve the robustness of presentation tracking, including the use of confusion networks to mitigate the effects of speech recognition errors, and the use of different weighting methods for semantic text retrieval.
CHAPTER 3

Exploratory Studies

To further understand the potential of presentation tracking to provide greater support for presentation delivery, I conducted two exploratory studies. In the first study, I examined the prevalence and forms of presentation notes, which can be an important source of content for presentation tracking. This study also helped inform the design of intelligent teleprompting systems that can reduce note reliance and improve presenters’ experience. In the second study, I investigated current practices of preparing and delivering dynamic presentations, which, according to my literature review, are not effectively supported in current presentation systems.

3.1 Exploratory Study on Notes Forms and Usage

To better understand current practices in the use of presentation notes, we reanalyzed 36 interviews conducted in previous studies regarding presentation practices [29, 112]. The interviewees (I1-36) were from a wide range of age, gender, nationality, education, and professions. We conducted thematic analysis [15] on the interview transcripts by first open coding, followed by clustering of relevant codes into common themes using the affinity diagraming method. In addition, we analyzed a corpus of presentation notes to gain a better understanding of the quantitative characteristics a typical presentation notes content. In the next subsections, we describe three themes that emerged from our analysis of interviews and present the results of the analysis of notes corpus.

1 This work was conducted in collaboration with Ha Trinh. It was published in [4].
3.1.1 Current Practices in Presentation Note Taking

Most of the interviewees (26 out of 36 participants) used notes for presentation authoring, rehearsal, or delivery. Their notes are in different forms, including handwritten or printed notes (7 participants), digital text documents (3 participants) or the slide notes sections in presentation software (11 participants). Some participants mentioned that they don’t use the note sections in slideware because it takes too much time to “get the whole thing set up” [I12] or didn’t know how to use them [I30]. Some presenters put notes on their actual presentation slides since they thought it might help the audience to follow the presentation [I9] or simply because of “being too lazy to rehearse and remember the bullet points” [I6], but some interviewees indicate that text-heavy slides distract the audience [I17, I21].

3.1.2 Application of Notes During Presentation Delivery

Participants used their notes to include optional content “that may not be spoken” [I9], presentation timing targets [I21], and most commonly as “a cheat sheet” [I13] to tell them about slide content. Many participants mentioned that forgetting the content is a common issue during presentations:

“I think when I was up there I got caught on the moment and lost track of some of the things I should have said.” [I21]

“I guess forgetting something, forgetting where I am is the challenge” [I22]

Presenters glance at their slides or notes to remember the content and flow of the presentation [I27]. Specifically, notes are used to remember verbal transitions between slides [I3], important take-home messages, and content that requires exact wordings, such as numbers and results [I21].

Interviewees said that one of the issues in using notes during delivery is the difficulty finding them during a presentation:
“try[ing] to find things in my note can take time and then there's a long pause in the presentation, which doesn't sound very good…” [I27].

Some interviewees also indicated that notes might encourage reading them “word by word” [I11] which can make the presentation disengaging [I1], less effective [I8] and less improvised [I5].

3.1.3 Best Practices and Requested Features

Interviewees suggested writing keywords and headlines [I6, I23] or highlighting the keywords in notes:

“I highlighted the key points... I can start talking about something else and then quickly look at the notes because I knew at this point I want to be looking at the red, so I'll keep look down to whatever red on my screen that was telling me what I was gonna be saying next.” [I25]

Another practice in note authoring that was mentioned is to divide the notes into segments related to bullet points on a slide [I22] or specific parts of slides [I27]. Segmentation can also separate the notes used for transition from other notes [I21]. Presenters also break up their notes to control the flow and pacing [I23].

Interviewees also suggested that a system that could automatically track their presentation and cue them could be helpful:

“If the note corresponding to what I was talking about [could] appear or something then it will be pretty cool” [I27]

Another issue raised is that displaying notes on computer screen is not optimal:
“[PowerPoint] has a feature where you can show notes but that's not really helpful if I'm standing in front of everybody and I'm very far away from the computer, but if you can make it like a big script thing rolling automatic.” [12]

3.1.4 Analyzing a Notes Corpus

To estimate a quantitative description of a typical presentation notes content, we analyzed a notes corpus containing 1.1 M words, gathered in a previous study [112], from 8000 PowerPoint presentations (21% of the presentations contained notes). The presentations covered a wide range of topics, from engineering to social sciences. Each presentation contained an average of 9 slides that have notes. Each slide note was structured into points corresponding to either paragraphs or bullet points. We extracted keywords in the notes using the method described in section 4.3.1. We calculated the statistics related to number of points, sentences, words and keywords from this corpus. Table 1 displays the results of this analysis.

Table 1. Statistics Gathered from the Notes Corpus

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Words per Slide</td>
<td>75.01</td>
<td>1</td>
<td>8059</td>
</tr>
<tr>
<td>Number of Points per Slide</td>
<td>3.64</td>
<td>1</td>
<td>176</td>
</tr>
<tr>
<td>Number of Sentences per Slide</td>
<td>5.41</td>
<td>1</td>
<td>431</td>
</tr>
<tr>
<td>Percent of Keywords per Slide (%)</td>
<td>55.56</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Number of Words per Point</td>
<td>20.59</td>
<td>1</td>
<td>435</td>
</tr>
<tr>
<td>Number of Sentences per Point</td>
<td>1.49</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>Percent of Keywords per Point (%)</td>
<td>58.72</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
3.2 Exploratory Study on Dynamic Presentations

We recruited 8 university professors (4 female, 4 male, P1-8) from computer science and history departments. All participants had extensive experience giving presentations in various settings, from conference presentations to classroom lectures, seminars, workshops and panel discussions. We conducted semi-structured interviews with open-ended questions about general presentation preparation and delivery practices, potential issues of linear presentations, and strategies for dynamic presentation support. Each interview lasted from 30 to 45 minutes.

The interviews were recorded, transcribed and coded using thematic analysis techniques [15]. Our initial open coding process resulted in 96 active codes capturing actions in data. Using affinity diagraming, we categorized these codes into the following three themes.

3.2.1 Need for Dynamic Presentations

Participants identified a variety of situations that require a more dynamic approach to presenting. The most cited scenario is post-presentation QA sessions. In these sessions, it is a common practice for presenters to navigate back to previously presented slides, or spontaneously show additional backup slides to support their responses:

"We will definitely build in backups and we encourage students to incorporate backups... During the question period, you feel like a million bucks if someone is like, ‘well what about this?’, and then you say, ‘well look at this!’" [P4]

Similar to QA sessions, unscripted presentations are also required for discussion-based settings such as “flipped classrooms” [P1] or “review classes” [P4]. In these scenarios, “inevitably it

---

1 This work was conducted in collaboration with Ha Trinh. It was published in [5].
involves a lot of shuffling. Someone just says, ‘I have a question about TCP’, and then OK, I have to go find that deck, and find that specific thing that you are talking about. So that’s two hours of jumping around in random places’” [P4].

In flexible and interactive settings such as classrooms or seminars, presenters also dynamically tailor their delivery based on their real-time evaluation of the audience’s level of understanding: “I’ll have a series of topics or modules that I want to cover in a class time, and some expand and some may contract depending on where the students are and their understanding and interest level” [P1]. In some cases, presenters also spontaneously select a subset of topics to present based on audience interest: “A lot of my presentations are play on adventure type things, where at some point I ask, ‘hey what do you guys want to talk about?’” [P8]. Most participants also mentioned the need for dynamically adapting their delivery due to time limits.

3.2.2 Preparing for Unscripted Interactions

Presenters often include backup slides at the end of their slide deck to prepare for QA sessions:

“If I am less familiar with a topic then I’m more likely to have some more backup slides...

I think the act of creating the slide is very important… at least it will put you at ease” [P2]

These backups can be new slides containing supporting material for anticipated audience questions, or existing slides that are “copied and pasted from other slide decks” [P8]. They can also be results of a timed rehearsal process, in which some slides are moved from the main slide deck to the backup section due to time constraints.

Presenters often write speaking notes for new presentations or when highly visual slides are used. Short, bullet point-style notes are generally preferred, although more verbose notes and specific wordings may also be used for important points when presenters really want to “get it across” [P3]. In certain settings such as history conferences, it is also common to recite full scripts during
presentations. Another approach is to make “those slides your notes” [P5], but this often results in text-heavy slides that are not desirable.

In addition to key speaking points, presenters also add “supplemental material” [P1] in their notes, such as more detailed explanations of a concept or additional information taken from an existing paper, which may be used when audiences ask related questions.

3.2.3 Nonlinear Slide Navigation

Most dynamic presentations require nonlinear access to relevant slides. The inability to quickly navigate to needed slides could cause significant disruption in the talk flow:

“It is disruptive if you can’t get there quickly, and anytime you muck with the computer you are risking messing yourself up... The last thing you want to do is wasting your time trying to find your slide” [P2]

This disruption and delay time in searching for slides could lead to a negative audience experience: “I have seen people go through 30 slides and finally go, ‘is this the slide you meant? or this slide?... it is a disaster” [P1]. Instead of skipping through slides in sequence, a more recommended method is to exit the presenter view and then directly select a slide thumbnail from the slide navigation pane. However, this can still lead to a long delay time, especially for big slide decks. Another tactic is to add slide numbers on slides and use them for direct keyboard navigation. However, this can still be problematic if slide numbers are misremembered. Due to these difficulties, presenters may opt to not show supporting slides during QA sessions, even though they might have spent time preparing them, unless they are confident that they know “where the slide is and it’s very easy to quickly go to slide” [P1].

To facilitate slide navigation, participants suggested having a dedicated “question mode that you could have flags for particular slides that you anticipated you needed” [P3], or indexing slides by
key phrases, so that “you know that a certain word would be phrase to a slide” [P4]. One participant envisioned a speech-based technology that could automatically display relevant slides based on the presenter’s speech:

“If there was a technology whether the screen would go black and within 5 seconds the slide I wanted would come up that would be great... If I could speak into my clicker and say, you know, ‘show me the slides that have this on it’ and maybe a few slides would pop up and I would click on the one” [P6].

In summary, findings from our interview analysis highlights the need for more intelligent support for seamless nonlinear slide navigation during dynamic presentation delivery.

3.3 Implications for Design

Informed by findings of our interview analysis, I propose four guidelines for the design of presentation tracking systems that support dynamic presentations and content-based feedback:

1. Support fast, nonlinear access to relevant supporting slides in real-time by dynamically suggesting the most related slides based on speech input from either presenters or audiences, in ways that minimize disruption to the presentation flow and avoid exposing the slide navigation process to the audience;

2. Enable tracking at sub-slide level through automatic identification of content coverage for each individual note segment within a slide’s notes;

3. Perform tracking in a manner independent of the presentation content ordering and exact content wordings, in order to allow for greater flexibility in presentation sequencing and speaker improvisations;

4. Accommodate a variety of presentation content and supporting note styles, from concise keywords, bullet points to verbose scripts.
3.4 Conclusion

Exploratory study on presentation notes usage showed that presentation notes are commonly used for including extra information in slides or remembering the content of the presentation. It was reported that navigating though notes might cause distraction during the presentation and thus a system that could automatically track the presenter’s progress in notes can be useful.

Analysis of interviews regarding dynamic presentations showed that question answering sessions are the most common scenario for non-linear presentations. Presenters prepared for these sessions by authoring extra slides that include supporting material related to potential audience questions. Interviewees also mentioned the need for better support systems for non-linear navigation between presentation slides.
CHAPTER 4

Tracking Framework

Based on the literature review and exploratory studies, I designed a tracking framework that tracks the coverage of presentation content during an oral presentation. In this chapter, I first present a requirement analysis for a presentation tracking system, before describing the designed framework. Finally, I present the implementation and evaluation of a tracking system based on the framework.

4.1 Requirement Analysis

Before designing the tracking framework, I determined the requirements for tracking and developed and evaluated a prototype system.

4.1.1 General Requirements

The tracking framework aims to determine the segments of presentation content that have been covered by the presenter in real-time. Based on findings from my literature review and exploratory studies, I gathered a set of general functionalities and conditions that need to be satisfied by the framework.

*Functional Requirements:*

- The tracking framework shall be able to identify covered sections of slide content based on speech input.
- The framework shall be able to perform tracking in both slide and sub-slide content segmentations, such as topics in a slide.
• The framework shall be able to semantically match the speech content and presentation content. It shall not depend on the order of presentation, sentence structure, or exact word forms.
• Tracking shall be performed in real-time during the presentation delivery.

Non-Functional Requirements:

• The tracking system should be accessible and integrated in common slide-ware.
• The tracking should be robust for different speakers and recording conditions
• The tracking should not depend on the subject of the presentation
• The tracking should be robust against improvisations and additions to the planned presentation content.
• The tracking should be performed with minimum amount of delay.

4.1.2 Prototype Development

To assess the feasibility of developing a tracking framework, estimate the accuracy, and obtain early feedback from users, I developed and evaluated a prototype tracking system. To perform semantic matching, the prototype system used an ASR module to spot keywords from slide notes in the presenter’s speech. The keywords were manually extracted from slide notes. In the following sections, I describe the different steps of performing presentation tracking in this prototype:

Step 1: Manual Slide Notes Processing

Slide notes were manually processed to extract content keywords based on their importance in the content. Notes for each presentation slide were segmented into topics, each containing multiple keywords. Common synonyms for each keyword were manually extracted using online thesaurus sources to remove dependency on exact words. Keyword which were used in multiple topics in
one slide were removed since they couldn’t be used for distinguishing a topic from other topics in
the slide.

*Step 2: Automatic Speech Recognition*

I used the open source Pocketsphinx ASR system to spot slide keywords in speech input. The
Generic US English acoustic model and a dictionary with 70,000 words were used for this purpose.
The pronunciations of out of vocabulary words from the slides were added to the dictionary
manually. The system returned the detected keyword at the end of each speech utterance.

*Step 3: Content Matching*

To track the presentation, the detected keywords were automatically matched against the slide note
keywords and their synonyms. Each topic was scored based on the ratio of the number of its
detected keywords over the total number of its keywords. The topics with scores higher than a
threshold were marked as covered. The threshold value of 0.3 was chosen empirically by running
the system on previous presentation recordings.

**4.1.3 Prototype Evaluation**

To evaluate the prototype tracking system, I integrated it into the DynamicDuo co-presentation
system and conducted a user study to evaluate the effectiveness of the system in improving
presenters’ experience.

*The DynamicDuo System:*

As previously mentioned in section 2.2.3, DynamicDuo [111] is a presentation delivery support
system that supports dual presentations between human and virtual presenters. Designed as a
PowerPoint add-in, DynamicDuo consists of three components: (1) a virtual agent that acts as a co-
presenter to delivery parts of the presentation; (2) a note authoring environment that enables the
presenter to create speaking notes for both the human and the co-presenter agent (Figure 8); (3) a
presentation delivery environment that displays the slide, notes and timing information, similar to the PowerPoint’s Presenter View.

The virtual agent is a life-sized, animated character developed using the Unity game engine. The agent communicates using synthetic speech and synchronized non-verbal behaviors, including facial expressions, gaze, and posture shifts. During presentation delivery, the agent is displayed on a secondary standing display next to the main slideshow display. The presenter navigates through the slides and triggers the agent’s speech using an IR remote clicker. When triggered, the agent presents her next assigned note segments. The presenter can manually pause and resume the agent’s speech at any time using the clicker.

A within-subjects user study showed that DynamicDuo led to significant improvements in speaker confidence for non-native English speakers, and in overall presentation quality for all presenters.

Figure 8. DynamicDuo note authoring environment: a) Notes for agent and human presenter. b) A note subsection assigned to the agent. c) Preview of the agent. Trinh et al. [111]
Integration in DynamicDuo:

Dynamic-Duo provides a strict method of interaction in which presenters need to plan and assign segments of notes to the agent and follow their planned sequence of turn-takings during the presentation delivery. To support less pre-planned presentations, I integrated the presentation tracking system into DynamicDuo to detect the segments of slide notes that were covered by the presenter. I added a speech-based interaction mode, in which pauses in presenter’s speech or specific voice commands could trigger the agent to present the next note topic that was not covered by the human presenter.

![Figure 9. A participant co-presenting with the virtual agent](image)

Evaluation Study:

In a within-subject two-sessions study, 15 participants were recruited to co-present with the agent using the original DynamicDuo system and the tracking enabled system. Participants presented two presentations with pre-made slide decks about Tigers and Lions (Figure 9). Each slide deck had 6
slides with detailed speaking notes containing 6-10 sentences. For each condition, participants had time to prepare and rehearse their presentations with the agent and then deliver their final video recorded presentations. They rated each co-presenter system using a 7-item, 7-point composite scale after their final presentations and at the end of both sessions they were interviewed about their experience.

**Evaluation Results:**

The composite rating of the co-presenter system with tracking (mean = 5.37, SD = 0.96) was higher than the non-tracking co-presenter system (mean = 5.09, SD = 1.2); however, there was no statistical significance (p=0.46). The recordings of the presentations were manually annotated to measure the accuracy of tracking system when a transition to the agent was triggered. The average accuracy was 71.5% (SD=24.6).

In post-study interviews, most of the participants mentioned that the tracking system made them feel more confident and reduced their anxiety since “it gives the feeling of someone covering you when you miss a point” [P1], “When you forget something, something is always there to back you up” [P7]. One participant reported that she asked the agent to present some of the sections that were originally assigned to the human presenter: “There were a lot of things to remember, so I gave some of them to Angela (the agent)” [P2]. This shows that the tracking system could support changes to the planned presentations during the delivery.

Some participants felt like the system was not accurate enough and it didn’t present the uncovered sections or repeated the sections that they had already presented. They reported that the system didn’t understand them well enough: “I would prefer if there was some way that she could read my mind. So, she could adapt to what I am thinking” [P9]. One participant mentioned that more practice with the system might improve the results: “I need to practice with her more. So, she can understand me better when I have to present.” [P2]
4.1.4 Summary

The prototype evaluation demonstrated the feasibility of a presentation tracking system and the potential utilization of tracking in presentation assistance applications. The prototype system achieved an adequate accuracy to motivate the development of a full tracking system. Based on early user feedback, tracking can improve the presenter’s experience and increase the speaker’s confidence by providing content related assistance.

The next steps in development included incorporating a more robust ASR system and automation of the slide content processing.

4.2 Design of the Framework

Based on my requirement analysis and the results of evaluation of the system prototype, I designed a presentation tracking framework. Improved upon the prototype system, the framework provides automatic slide content processing and utilizes a more robust method for speech content processing.

Following are the details of the tracking framework and the results of evaluation of the framework accuracy (published in [3]).

The tracking framework consists of two main units: offline slide content processing and real-time speech to content matching. Before the presentation, slide contents keywords are automatically extracted and weighted. At runtime, the framework detects these keywords in the ASR confusion network output and based on the detected keywords and their weights, slide contents are scored. Figure 10 shows the overall architecture of the framework.
4.2.1 Slide Content Processing

Slide content are extracted from each presentation slide and processed. Slide contents may include slide notes, slide textual content, and meta information about slide visual elements.

Content Segmentation:

The slide content is segmented into sections to make tracking more refined. If the sections are too small (such as at word level) overfitting might happen and small deviations from the notes could result in false negative results. Some of the methods for dividing content into sections include sentence segmentation, semantic topic segmentation, or manual topic segmentation by the presenter.

Keyword Extraction and Expansion:

To perform semantic matching, the framework extracts important words in each segment. Information retrieval and keyword extraction methods mentioned in section 2.4 can be used for this purpose. To remove reliance on exact words, keywords are lemmatized and their synonyms are extracted.
**Keyword Weight Assignment:**

To consider the importance of each keyword in slide content, the framework assigns weights to keywords using two methods: *tf.idf* and semantic similarity.

**Term frequency-inverse document frequency (tf.idf):** As mentioned in section 2.4, *tf.idf* is a common information retrieval method used for scoring words based on their importance to a document in a collection of documents. To use *tf.idf* in presentation tracking, each slide segment or slide can be considered as a document and the slide or the presentation as a whole can be considered as the collection of documents. *tf.idf* is calculated using the following equation:

\[
tf\cdot idf(w, S) = tf(w, S) idf(w)
\]

\[
idf(w) = \log\left(\frac{|D|}{|\{\forall S \in D \mid w \in S\}|}\right)
\]  

(1)

Where \(tf(w, s)\) is the frequency of word \(w\) in document \(S\), and \(D\) is the set of all documents. \(idf(w)\) is the logarithmically scaled inverse fraction of the documents that contain the word \(w\), which corresponds to how common the word \(w\) is in the collection of documents \(D\). The reasoning behind \(idf\) is that the words used in several documents have low specificity for each of those documents. Therefore, compared to a unique word, they are less useful for identifying the document containing them.

One issue in using *tf.idf* is that it gives high scores to words that, although specific to a document, are not semantically important in that document. These words can be omitted during the presentation without affecting the general meaning but due to their high *tf.idf* scores their omission might result in false negative results. To fix this issue, I combine *tf.idf* weights with weights assigned based on semantic specificity of words, which are described next.
**Semantic specificity weights:** Inspired by [62], I argue that an important keyword should be more semantically relevant to the document containing it compared to the other documents in the corpus.

I model this concept using the semantic specificity ratio score $ssr$:

$$ssr(w) = \frac{local\_similarity(w, S)}{global\_similarity(w, D)}$$  

(2)

For a keyword $w$ from document $S$, $local\_similarity(w, S)$ is the similarity of $w$ to other keywords in document $S$, and $global\_similarity(w, D)$ is the similarity of keyword $w$ to the keywords in the documents in corpus $D$ other than document $S$.

To measure the semantic similarity, the framework uses word embedding. Word vectors representing more semantically similar words have smaller Euclidean distance and bigger cosine similarity.

To calculate the cosine similarity between a word and the word set containing that word, I used the average cosine similarity between the word and other words in the word set:

$$\forall w_i \in S \quad s_c(w_i, S) = \frac{\sum_{w_j \in S} csim(w_i, w_j)}{|S| - 1}$$

(3)

$$csim(w_i, w_j) = \frac{\sum_{k=0}^{n} v(w_i)_k \cdot v(w_j)_k}{\sqrt{\sum_{k=0}^{n} v(w_i)_k^2} \cdot \sqrt{\sum_{k=0}^{n} v(w_j)_k^2}}$$

To calculate the similarity using the Euclidean distance, I used a form of Closeness Centrality measure, which has been used in graph based key phrase extraction [14]:

$$\forall w_i \in S \quad s_d(w_i, S) = \frac{|S| - 1}{\sum_{w_j \in S} dist(w_i, w_j)}$$

(4)
\[
\text{dist}(w_i, w_j) = \sqrt{\sum_{k=0}^{n} (v(w_i)_k - v(w_j)_k)^2}
\]

The value of \( n \) in equations 3 and 4 is equal to the number of dimensions of the word vectors.

Finally, the semantic specificity ratio in equation 2 is calculated using the cosine similarity or Euclidean distance:

\[
\text{ssr}(w, S) = \frac{\text{csim}(w, S)}{\text{csim}(w, D)} \text{ or } \frac{\text{dist}(w, S)}{\text{dist}(w, D)} \quad (5)
\]

Where \( S \) is the set of words in the document \( S \) containing \( w \) and \( D \) is the set of words in other documents in corpus \( D \). Semantic specificity ratios are multiplied by \( \text{tf.idf} \) weights and these combined weights are normalized by dividing the weight of each keyword by the sum of weights of all of the keywords in the segment:

\[
\forall w_i \in S \quad \text{ssw}(w_i, S) = \frac{\text{ssr}(w_i, S)\text{tf.idf}(w_i, S)}{\sum_{w \in S} \text{ssr}(w, S)\text{tf.idf}(w, S)} \quad (6)
\]

**Sample weighting scenario:**

To clarify the scoring process, I present a sample keyword scoring scenario. Figure 11 shows the notes for a sample slide with 7 segments. The keywords are in bold.

Table 2 shows the weights for keywords of segment 3. The table is ordered by normalized similarity weights. We can see that \( \text{tf.idf} \) discards the word “tiger” since it is used in all segments. Semantic specificity weights give higher weights to “orange”, “black” and “color” since they are semantically close to each other and represent the main segment concept. The word “common” is weighted highly by \( \text{tf.idf} \) since it is only used in this segment but it has the lowest weight in similarity ratios.
Normalizing the weights results in lowering the final weight for “common”. This effect is more evident in $ssw_c$, which uses the cosine similarity weight.

1. The tiger is the largest cat species.
2. An adult male wild tiger can reach a total body length of up to 11.5 feet, and weigh up to 850 pounds.
3. The most common color of tigers is orange, with black stripes.
4. But each tiger has a unique stripe pattern, much like our fingerprints.
5. We have also seen some color variations, with white, black, golden tabby, and blue tigers.
6. The current population of wild tigers is estimated to be about 3200 individuals.
7. There are 10 recognized tiger subspecies, but four of them are considered extinct.

Figure 11. Sample slide notes with keyword candidates in bold.

Table 2. Keywords and weights from segment 3 of sample notes

<table>
<thead>
<tr>
<th></th>
<th>$tf.idf$</th>
<th>$ssr_c$</th>
<th>$ssr_d$</th>
<th>$ssw_c$</th>
<th>$ssw_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>orange</td>
<td>0.845</td>
<td>1.113</td>
<td>1.3</td>
<td>0.293</td>
<td>0.284</td>
</tr>
<tr>
<td>common</td>
<td>0.845</td>
<td>0.668</td>
<td>1.106</td>
<td>0.176</td>
<td>0.242</td>
</tr>
<tr>
<td>color</td>
<td>0.477</td>
<td>1.147</td>
<td>1.295</td>
<td>0.170</td>
<td>0.160</td>
</tr>
<tr>
<td>black</td>
<td>0.477</td>
<td>1.086</td>
<td>1.281</td>
<td>0.161</td>
<td>0.158</td>
</tr>
<tr>
<td>stripes</td>
<td>0.477</td>
<td>1.347</td>
<td>1.267</td>
<td>0.200</td>
<td>0.156</td>
</tr>
<tr>
<td>tiger</td>
<td>0</td>
<td>0.887</td>
<td>1.15</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

4.2.2 Real-Time Content Matching

In this unit, presenter’s speech content is semantically matched against the slide content to identify the segments covered by the presenter.

*Keyword Spotting:*

To spot the slide content keywords, the framework matches the ASR output against the keyword lemmas and their synonyms. To decrease the dependency on the accuracy of the ASR system and
increase the robustness of the tracking, the framework uses confusion networks. As mentioned in section 2.3.2, ASR output in the form confusion network provides much richer information compared to the best hypothesis output. Some ASR systems provide confusion networks as an output option but if not available the ASR lattice output can be decoded to generate the confusion networks graph [67].

The confusion network contains alternative word hypotheses in each time frame ordered by their acoustic confidence score. The framework iterates through this ordered list and compares each hypothesis with the slide content keyword lemmas and their synonyms. If there is a match, other alternative hypotheses in that time slot are discarded and the matched keyword is tagged as spotted in all segments containing that keyword. Each keyword can only be spotted once in a slide. Figure 12 shows a sample keyword spotting scenario.

**ASR output for a time frame sorted based on acoustic confidence level**

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Alternative</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.62</td>
<td>variation</td>
<td>fluctuation</td>
</tr>
<tr>
<td>0.31</td>
<td>alteration</td>
<td>change, modification</td>
</tr>
<tr>
<td>0.06</td>
<td>operation</td>
<td></td>
</tr>
</tbody>
</table>

**Sample Content keywords**

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>photo</td>
<td>photograph</td>
</tr>
<tr>
<td>color</td>
<td>colour</td>
</tr>
<tr>
<td>adjustment</td>
<td>modification</td>
</tr>
<tr>
<td>crucial</td>
<td>important</td>
</tr>
</tbody>
</table>

“modification” is matched

“adjustment” is spotted

**Figure 12. A sample keyword spotting scenario.**

*Content Scoring:*

Each time a keyword is spotted the score for the segments containing it will increase by the amount equal to the weight for that keyword. Therefore:
\[ s(S) = \sum_{w \in S \cap R} ssw(w, S) \]

Where \( s(S) \) is the score for the segment \( S \), \( R \) is the set of words spotted in speech, and \( ssw(w, S) \) is the weight for keyword \( w \) in segment \( S \).

This score can be used for different applications of presentation tracking. For example, for content coverage, segments with scores higher than a threshold are tagged as covered and for determining the content segment that is most relevant to the speech content, we can look at the segment with the highest score.

**4.3 Implementation and Evaluation of a Tracking System**

I implemented and evaluated a presentation tracking system, which was developed based on the tracking framework. This system uses common techniques and off-the-shelf tools for speech recognition and natural language processing. In the next sections, I describe the technical details in the implementation of the system.

**4.3.1 Implementation Details**

*Keyword Extraction:*

The system uses Stanford CoreNLP tools [68] for segmenting the slide notes into sentences, and performing the part of speech tagging. It removes stop words, punctuation marks and symbols and replaces numbers with their word representations. Also, adverbs are removed. The remaining words are lemmatized to remove dependency on exact word forms and used as keywords.

*Keyword Expansion:*

WordNet [74] is used to extract the synonyms for each keyword. To do so, the most common synset for each word is retrieved and the words in that synset are extracted as that word’s synonyms. If a
word is a synonym for multiple words, the word in the synset with highest tagged frequency in WordNet is chosen. WordNet stemming is used in addition to CoreNLP lemmatization for comparative and superlative adjectives. Extracted keywords and their synonyms are stored in a table.

.Keyword Weighting:

Keywords are weighted based on semantic specificity and \textit{tf.idf} measures. For semantic specificity weighing the system uses a pre-trained GloVe vector representation [84] with 1.9 million uncased words and vectors with 300 elements. It was trained using 42 billion tokens of web data from Common Crawl.

\textit{Speech Recognition:}

Automatic speech recognition is performed using IBM’s Watson cloud-based service\textsuperscript{1}. This service provides both \textit{n}-best transcripts and confusion networks. The tracking system discards confusion network alternative hypotheses with the level of confidence lower than 0.01. This threshold was chosen empirically.

\textbf{4.3.2 Evaluation Experiments}

I evaluated the performance of the tracking system using the corpus of presentation recordings gathered during the prototype evaluation. The evaluation corpus included 30 videotaped presentations delivered by 15 speakers. Each presentation contained 5 slides, with an average length of 5 minutes. Recordings were split for each slide, resulting in 150 slide presentation recordings.

\footnote{\url{https://www.ibm.com/watson/services/speech-to-text/}}
To simulate real-time tracking conditions, each slide presentation recording was segmented into 3 sections in a semi-random manner. This was done by detecting the 2 longest pauses in speech and splitting the recording around them. I also made sure that each segment is at least 6 seconds long. Using this segmentation method, each speech segment might cover a random number of sentences from zero to the total number of sentences in the slide. A few recordings were too short to be split and were discarded. This process resulted in a total of 426 audio files.

Each audio file was manually annotated for content coverage by a human annotator. The annotator was instructed to subjectively tag a sentence as covered if she found that the main points of the sentence were covered in sufficient detail. 100 recordings were randomly chosen and annotated by another annotator to check the inter-rater agreement. The Cohen’s Kappa coefficient was 0.84, which indicates a high degree of agreement.

To evaluate the system, the speech samples were automatically annotated for content coverage using different tracking methods and thresholds. Manual and automatic annotations were compared and precision, recall, and f-score measures were calculated for each method.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>Using 1-best ASR output</td>
</tr>
<tr>
<td>synons</td>
<td>baseline + synonyms</td>
</tr>
<tr>
<td>words</td>
<td>confusion network ASR output + synonyms</td>
</tr>
<tr>
<td>tfidf</td>
<td>words method + tf.idf score</td>
</tr>
<tr>
<td>cosine</td>
<td>tfidf method + cosine similarity weighting</td>
</tr>
<tr>
<td>distance</td>
<td>tfidf method + Euclidean distance weighting</td>
</tr>
</tbody>
</table>

I used a tracking method similar to [81] as the baseline. In this method, the 1-best ASR results were matched against the slide notes. Note sentences were scored based on the ratio of the spotted keywords to total number of the keywords in sentence. I evaluated the tracking using different ASR
outputs, keyword extraction methods, and keyword weighting methods. The systems were evaluated using different threshold values for content coverage. Table 3 lists the evaluated tracking systems and their reference names in the next section.

4.3.3 Evaluation Results

Figure 13 shows the precision-recall curves for 3 tracking systems with content coverage threshold changing from 0 to 1. It also includes the curves for the highest F-score values for distance and baseline methods. Precision, recall and F-score values increased in both words and distance methods compared to the baseline method. Increasing the threshold generally results in lower recall but higher precision values. Threshold values between 0.2 and 0.3 led to the best F-scores for all methods.

![Figure 13. Precision-Recall curves for different tracking methods](image-url)
Table 4 shows the values of precision, recall and F-score for each system using threshold values that were optimized for highest F-score. An approach that randomly tags sentences as covered is also included in the table for comparison.

**Table 4. Evaluation measures using thresholds optimized for best F-score**

<table>
<thead>
<tr>
<th>Method Name</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F-score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>random</td>
<td>25.53</td>
<td>53.44</td>
<td>34.55</td>
</tr>
<tr>
<td>baseline</td>
<td>68.13</td>
<td>65.11</td>
<td>66.59</td>
</tr>
<tr>
<td>synons</td>
<td>70.32</td>
<td>71.06</td>
<td>70.69</td>
</tr>
<tr>
<td>words</td>
<td>74.30</td>
<td>76.55</td>
<td>75.40</td>
</tr>
<tr>
<td>tfidf</td>
<td>76.24</td>
<td>76.78</td>
<td>76.51</td>
</tr>
<tr>
<td>cosine</td>
<td><strong>79.70</strong></td>
<td>73.75</td>
<td><strong>76.61</strong></td>
</tr>
<tr>
<td>distance</td>
<td>72.00</td>
<td><strong>82.50</strong></td>
<td><strong>76.89</strong></td>
</tr>
</tbody>
</table>

We can see the positive effect of using synonyms in 4.1% F-score improvement from baseline to synons. Using confusion network instead of the best hypothesis increased the F-score by 4.71%. Using keyword weighting improves the F-score and semantic weighting methods perform better than only using tf.idf weighting. Using the Euclidean distance weighting results in the best recall value and the cosine method has the best precision. Depending on the application requirement, we can choose between these two semantic weighting methods but in general distance method has the best F-score of 76.89%.

**4.4 Conclusion**

In this chapter, I described a presentation tracking framework that can track the current state of the oral presentation by semantic matching of presenter’s speech content and presentation content. To do so, the system extracts keywords from the slide content, and scores them based on tf.idf and
semantic specificity. The presentation content is scored based on detected keywords in ASR confusion network output.

I also described a tracking system based on this framework. An evaluation study showed that the performance of the tracking system improved by using confusion networks, semantic matching and keyword weighting. In the best system, the F-score increased 10.3% compared to a baseline method.

In the following chapters, I describe user studies in which I demonstrated the effectiveness of the tracking framework in different presentation assistance applications.
CHAPTER 5

IntelliPrompter System

In this chapter, I examine one of the applications of presentation tracking during presentation delivery. I present IntelliPrompter, a novel extension to slideware that provides an intelligent fully-automated teleprompter that listens to a speaker’s presentation and cues them regarding what should be talked about next during their presentation. The system allows flexibility in the sequence of slides presented, the order of topics within a slide, and the wording the speaker uses, allowing for much greater ability for speakers to improvise. Automatic teleprompting through presentation tracking can lessen the fear of forgetting important points and therefore improve the presenter’s experience during the presentation delivery. I also present the results of investigating the use of a head-mounted display for teleprompter cues to improve eye contact with the audience compared to standard “Presenter view” on a laptop.

This chapter includes: 1) the design of a system to automatically track speaker progress through their slides and notes; 2) the design of two different display modalities for cueing speakers in real time; and 3) results of a summative evaluation comparing the system and two display modalities to standard slideware practice. The contents of this chapter are published in [4].

---

1 This study was conducted in collaboration with Ha Trinh, Harriet Fell, and Timothy Bickmore
5.1 Design of the Note Display Systems

Based on our findings from the exploratory study described in Chapter 3 we designed a note display system integrated within PowerPoint. The system listens to a presenter’s speech during his or her presentation, automatically tracks how much of the current slide’s notes have been covered, and displays cues to the speaker about what should come next in their presentation. The system does not require the presenter to follow the order of their notes within each slide, and does not rely on exact words or sentence structures, allowing it to support more flexible and spontaneous presentations. The interface highlights the keywords in the segment of the notes that are most likely to be presented next.

A head-mounted display can provide unique affordances for displaying information to presenters, allowing them roam away from fixed teleprompters during their speech, and allowing them to surreptitiously refer to their notes without breaking eye contact with the audience. Accordingly, we explored both a Google Glass-based version and a fixed computer monitor version of our interface for evaluation.

5.1.1 Authoring Environment

Our system provides a note authoring interface, implemented as an add-in for Microsoft PowerPoint 2016, which enables the presenter to prepare speaking notes for each slide (Figure 14).

Figure 14. Authoring Environment (Left) and Presenter View (Right)
This note authoring tool allows the presenter to divide the speaking notes of each slide into a series of segments which we call topics. The presenter can enter notes, specifying what they plan to say for each topic. As we heard in our interviews, organizing slide notes into topics is a common and helpful practice, and in addition it can provide a means of segmenting the notes into smaller sections for the purpose of tracking.

5.1.2 Presentation Mode

During a presentation, the system tracks the presenter’s speech to determine which topics have been covered on each slide. At each moment, one topic is highlighted and displayed at the top of the notes pane. The keywords in that topic are highlighted in a different color and a bigger font so as to be distinguished from rest of the text. The order of topics is determined based on the following process:

1- At the beginning of the presentation of each slide, note topics are displayed in the originally-authored order, and the first topic is highlighted.

2- During the presentation, anytime the system tags a topic as “covered” (i.e., the speaker has mentioned it), it adds a tick mark next to that topic and highlights the next uncovered topic in the topics sequence.

3- The notes scroll up to put the new highlighted topic at the top of the notes pane. The speed of scrolling was determined empirically to display the next topic as soon as possible while avoiding distraction.

4- If the user does not follow their originally-authored topic order during the presentation, the uncovered topics originally preceding the just-covered topic are moved to the bottom of the list of topics.
As mentioned before, we designed the system with two display modalities:

**Computer Screen Display:** In this version, we provide a view similar to PowerPoint’s Presenter View, which displays slide notes in a notes pane, along with the slide and timing information (Figure 14).

**Google Glass Display:** Only the notes for the next topic are displayed, with keywords highlighted as above. Because of the small size of the display, we use a font size for which the display can show at least 20 words (the average notes topic length).

### 5.2 Evaluation of Note Display Systems

We conducted a user study comparing our two dynamic, speech-based note display interfaces against a standard, static note display system that does not perform automatic presentation tracking (control condition). Our aims in this study were to examine the effectiveness of our speech-based note display systems in improving the overall experiences of presenters and audiences, and to compare two modalities (Screen-based vs. Google Glass) for displaying speaking notes.
5.2.1 Note Display Interfaces

We compared the three following note display interfaces:

*Static Screen Display (Control):* Similar to conventional slideware, the system displays the Presenter View on a secondary 15-inch computer screen showing the current slide and all speaking notes for that slide. During delivery, the presenter can use a remote control to linearly navigate to the topic that he/she wants to say next. The selected topic will be highlighted and moved to the top of the note pane for easy readability.

*Dynamic Screen Display:* Similar to the Control condition, the system displays all speaking notes for the current slide on a computer screen. During delivery, the system automatically tracks the content coverage based on the presenter’s speech and dynamically adjusts the note display (as described in Section 4.2). As a backup strategy, the presenter can also use the remote control to manually navigate to the next topic if needed.

*Google Glass:* As in the Dynamic Screen Display condition, the system tracks content coverage and dynamically displays the next note topic on the Google Glass screen, and allows the user to advance the topic using a remote control. A Presenter View is also provided on a separate computer screen, but only the current slide without any notes is displayed.

5.2.2 Procedure

We asked each participant to rehearse and deliver three 3-minute presentations on comparable topics (Italy’s tourist attractions, Italian art, and Italian cuisine) in English. Each presentation contained 3 pre-made slides. Based on our previous note corpus analysis, we supplemented each slide with an average of 79-word speaking notes spanning 4-7 sentences (mean = 5.7), divided into 4-5 topics (mean = 4.4). Each topic had a mean length of 18 words, 55 percent of which were identified as keywords on average.
The study was a within-subjects, single-session design with three conditions: *Dynamic Screen Display* vs. *Google Glass* vs. *Control*. Each session lasted approximately 90 minutes. The ordering of the conditions and the slide decks were randomly assigned and counterbalanced.

![Figure 16. A participant presenting with Google Glass](image)

At the beginning of the session, we introduced participants to the scenario of presenting three pre-made slide decks using different note display systems. We instructed them to cover all the main points in the notes, emphasized that they could rephrase the notes during their presentation. For each condition, we gave participants 5 minutes to review the notes, before performing one round of spoken rehearsal that lasted approximately 5 minutes. At the beginning of each rehearsal, we gave participants brief instructions on how to use one of the three note display systems. Following each rehearsal, we asked participants to give their final, videotaped presentation in front of the experimenter. We concluded the session with a semi-structured interview, probing for comparisons of the three systems and their effects on the experience of the presenters.
To maintain consistency across all conditions, we asked participants to wear the Google Glass in all rehearsal and delivery rounds, as shown in Figure 16. During the rehearsal and delivery, the experimentation room was set up with a 50-inch display monitor for slide projection, a tripod-mounted microphone for recording the presenter’s speech, and a secondary 15-inch computer screen displaying the Presenter View.

5.2.3 Presenter Participants

We recruited 31 students and professionals (15 female, 17 male, ages 21-35, mean 25) with backgrounds in science, technology, engineering, business and management. However, one participant had to be dropped from the study due to a technical issue. Of the remaining 30 participants, 1 was categorized as a low competence public speaker, 18 were categorized as moderate competence public speakers, and 11 had high competence according to the Self-Perceived Communication Competence Scale [69]. Participants were compensated for their participation.

5.2.4 Presenter Measures

Presenters’ experience with the note display systems were assessed using the following self-reported measures:

![Figure 17. Absolute ratings of three note display systems (* indicates statistical significances in pairwise comparisons)](image_url)

Figure 17. Absolute ratings of three note display systems (* indicates statistical significances in pairwise comparisons)
Absolute Usability Rating of Note Display System: Assessed in each condition after delivering each presentation, using a 7-item, 7-point composite scale, as shown in Figure 17.

Relative Usability Rating of Note Display Systems: Assessed after delivering all three presentations, ranking the three note display systems from 1 (Best) to 3 (Worst) on the same 7 criteria included in the Absolute Usability Rating.

Comfort of Wearing Google Glass: Assessed after delivering the presentation in the Google Glass condition, using a single-item, 7-point questionnaire.

Self-perceived Rating of Presentation Quality: Assessed in each condition after delivering each presentation, using a 7-item, 7-point composite scale evaluating the presenter’s self-perception of the engagingness, understandability, nervousness, excitement, entertainingness, competency and overall satisfaction of their presentation.

State Anxiety: Assessed before each presentation using the validated 20-item State Anxiety questionnaire [105].

Speaker Confidence: Assessed at intake and after each presentation using the 30-item Personal Report of Confidence as a Speaker questionnaire [83].

5.2.5 Presenter Quantitative Results

Absolute Usability Rating of Note Display System:

Figure 17 shows the results of the presenters’ absolute ratings of the three note display systems. Results of a repeated-measures ANOVA test on the composite rating (Cronbach’s α=0.89) showed significant differences between the three note display conditions ($F_{2,58}=8.023, p=.001$, partial $\eta^2=.217$). Post-hoc analysis using t-tests with Bonferroni correction showed that the Dynamic Screen Display was rated significantly higher than the Control condition ($p = .01$) and the Google
Glass (p = .001). There was no significant difference in the composite usability rating between the Control and the Google Glass conditions.

Results of Friedman tests on individual scale items revealed significant effects of the note display condition on ease of use ($\chi^2(2) = 13.83$, $p = .001$), reliability ($\chi^2(2) = 9.64$, $p = .008$), helpfulness ($\chi^2(2) = 6.07$, $p = .048$), distraction ($\chi^2(2) = 12.02$, $p = .002$), satisfaction ($\chi^2(2) = 6.54$, $p = .038$), and desire to continue using the system in future presentations ($\chi^2(2) = 8.11$, $p = .017$). Post-hoc analysis using Wilcoxon signed-rank tests with Benjamini-Hochberg correction showed that the Dynamic Screen Display was rated significantly higher than the Control condition in terms of ease of use ($Z = -2.87$, $p = .006$), helpfulness ($Z = -3.23$, $p = .003$), satisfaction ($Z = -2.79$, $p = .008$), and desire to continue using the system in future presentations ($Z = -2.32$, $p = .03$). The Dynamic Screen Display was also rated significantly better than the Google Glass condition on ease of use ($Z = -3.34$, $p = .003$), reliability ($Z = -3.34$, $p = .003$), helpfulness ($Z = -2.22$, $p = .039$), distraction ($Z = -3.19$, $p = .003$), satisfaction ($Z = -2.83$, $p = .008$), and desire to continue using the system in future presentations ($Z = -3.06$, $p = .006$). Compared to the Control condition, the Google Glass condition was rated significantly more distracting ($Z = -2.13$, $p = .05$). No other significant differences were found in other pairwise comparisons after correction.

Relative Usability Rating of Note Display System:

Results of Friedman tests revealed significant effects of the note display condition on helpfulness ($\chi^2(2) = 6.07$, $p = .048$), distraction ($\chi^2(2) = 6.87$, $p = .032$), and desire to continue using the system in future presentations ($\chi^2(2) = 7.48$, $p = .024$). There were no statistical differences in the relative rankings of the three conditions in terms of ease of use ($p = .061$), reliability ($p = .177$), satisfaction ($p = .393$) and how obvious it was to see what to say next ($p = .072$).

Post-hoc analysis using Wilcoxon signed-rank tests with Benjamini-Hochberg correction showed that the participants expressed significantly higher desire to continue using the system in future
presentations for the Dynamic Screen Display ($Z = -2.14, p = .048$) and the Google Glass ($Z = -2.36, p = .048$), compared to the Control condition. The Dynamic Screen Display was also ranked as significantly less distracting than the Google Glass ($Z = -2.36, p = .012$). No other significant differences were found in other pairwise comparisons after correction.

*Comfort of Wearing Google Glass:*

Participants were generally comfortable wearing the Google Glass during presentation delivery (mean = 5.37, SD =1.97), and this was significantly higher than a ‘neutral’ score of 4 (Wilcoxon signed-rank test, $p = .002$).

*Self-perceived Rating of Presentation Quality:*

Results of a repeated-measures ANOVA test on the composite rating of presentation quality (Cronbach’s $\alpha=0.77$) showed no significant differences between the three conditions.

*Speaker Confidence and State Anxiety:*

Results of a repeated-measures ANOVA test showed a significant interaction effect of *Condition* *Gender* on speaker confidence ($F_{2,56}=3.796, p=.028$, partial $\eta^2=.119$). Female participants reported significantly higher speaker confidence in the *Dynamic Screen Display* condition compared to the *Control* condition ($p=.013$). The difference in speaker confidence change also approached significance between the *Google Glass* and *Control* conditions ($p=.054$), in favor of the *Google Glass*. There were no significant differences between the *Dynamic Screen Display* and the *Google Glass* for female participants, and between the three conditions for male participants.

Results of a repeated-measures ANOVA test showed no significance effects of the condition on State Anxiety.
In summary, the presenters’ quantitative results demonstrated a clear positive effect of the Dynamic Screen Display to improve their experience during presentation delivery, compared to both the Control and the Google Glass conditions. In contrast, there were no major differences between the Google Glass and the Control condition, except that participants were more willing to continue using the Google Glass for future presentations even though it was perceived as more distracting.

5.2.6 Judge Evaluation of Presentation Quality

To evaluate an audience’s perception of the quality of presentations delivered using the three note display systems, we asked independent judges to review and evaluate the videotaped presentations. We recruited 30 judges (20 male, 10 female, ages 22-60, mean 27), who were students, researchers, lecturers and other professionals with diverse educational backgrounds and varying levels of presentation experience.

Each judging session lasted approximately 60 minutes. In each session, we asked a judge to watch two sets of three presentations each from two presenter participants, resulting in 6 presentations in total. The ordering of the presentations was randomly assigned and counterbalanced across the judges.

Judges were asked to complete the following questionnaires:

*Absolute Rating of Presentation Quality:* Completed after watching each presentation, using a 7-item, 7-point composite scale assessing the engagingness, understandability, novelty, excitement, entertainingness, overall quality and desire to continue seeing similar presentations.

*Audience Perception of Presenters:* Completed after watching each presentation, using a 7-item, 7-point composite scale assessing the presenter’s competency, engagingness, nervousness, understandability, excitement, entertainingness and overall satisfaction.
Relative Rating of Presentations: Completed after watching a pair of presentations from the same presenter. Each pair was compared on 6 criteria adopted from [21], including: organization, content coverage, note reliance, speech quality, timing and pacing, and overall quality. Each criterion was judged on a 4-point ordinal scale of “no difference”, “slight difference”, “moderate difference”, and “substantial difference”, with an indication of the superior presentation, if any.

Judge Rating Results:

Absolute Rating of Presentations: Results of a repeated-measures ANOVA test with Greenhouse-Geisser correction showed a significant effect of study condition (F_{1.58,93.22}=7.48, p=.002, partial η^2=.112) on the composite rating of presentation quality (Cronbach’s α=0.95). Judges rated presentations delivered using the Dynamic Screen Display (p=.014) and the Control condition (p=.004) significantly higher compared to the Google Glass condition. There was no significance difference between the Dynamic Screen Display and the Control conditions.

Audience Perception of Presenters: Results of a repeated-measures ANOVA test with Greenhouse-Geisser correction showed no significant effect of study condition on the composite rating (Cronbach’s α=0.83).

Relative Rating of Presentations: Results of Wilcoxon signed-rank tests showed that judges rated the speech quality of presentations to be significantly better in the Dynamic Screen Display (p=.004) and the Control condition (p=.006), compared to the Google Glass condition. No significant differences were found for the other criteria.

In summary, judges rated presentations delivered with the Google Glass condition as having significantly worse presentation and speech quality compared to the other two conditions.

5.2.7 Objective Measures

We calculated four objective measures to further evaluate the effectiveness of the system.
Accuracy of the Presentation Tracking System:

To calculate the accuracy of the system, a research assistant annotated the presentation recordings to check for content coverage. The annotator was instructed to listen to audio recordings of presentations for each slide and tag the time at which each topic in the notes were covered in sufficient details during the presentations. 10% of the recordings (27 out of 270) were randomly chosen and annotated by another annotator to check the inter-rater agreement. We considered the annotations for each annotator as equal only if they both tagged the same sequence of topics for a slide. Based on this, the Cohen’s Kappa coefficient was 0.73, which indicates a moderate degree of agreement.

To evaluate the system performance against human annotations, we extracted the times at which each topic was tagged as covered from system logs. Each system tag for a topic was marked as correct only if it occurred within the time span between the given topic and its previous topic as indicated in human annotations. The average accuracy of the system for all of the presentations in Dynamic Screen Display and Google Glass conditions was 75.65% (SD=16.69).

Results of a repeated-measures ANOVA test showed a significant effect of Gender on the accuracy of the system (F_{1,28}=8.96, p=.006, partial η²=.243). The system had significantly higher accuracy for female participants (mean=82.46%, SD=3.11) compared to the male participants (mean=69.69%, SD=2.91). This can potentially explain the significant interaction effect of Condition*Gender on speaker confidence that we reported in section 5.3.5.

Results of Pearson’s correlation test showed no significant correlation between the accuracy of the system and the speaker’s confidence in Dynamic Screen Display (r = -.098, n = 30, p = .608) or Google Glass (r = .009, n = 30, p = .961) conditions. There was no significant correlation between the accuracy and the presenters’ composite absolute ratings of the system in Dynamic Screen Display (r = .246, n = 30, p = .189) or Google Glass (r = -.096, n = 30, p = .612) conditions, but in
Dynamic Screen Display condition there were significant correlations between the accuracy of the system and the reliability \( r=0.380, n=30, p=0.038 \) and helpfulness \( r=0.469, n=30, p=0.009 \) of the system.

Content Coverage:

Using the same human annotations, we also calculated the content coverage for each presentation. Average content coverage was 91.06\% (SD=2.82) in the Control condition, 90.88\% (SD=2.77) in the Dynamic Screen Display condition, and 85.62\% (SD=3.4) in the Google Glass condition. Results of a repeated-measures ANOVA test showed no significant differences between the three conditions \( F_{2,58}=2.84, p=0.069 \).

Note Reliance:

A research assistant annotated the presentation video recordings to check for note reliance. The annotator was instructed to measure the time period that each speaker spent looking at the notes. For Google Glass condition, we counted the time presenter looking at the Glass display. This measured time was divided by total duration of presentation to calculate the normalized note reliance measure. 10\% of the recordings (9 out of 90) were randomly chosen and annotated by another annotator to check the Intraclass Correlation Coefficient which was 0.984.

Average Note Reliance was 44.26\% (SD=4.16) in the Control condition, 44.12\% (SD=3.79) in the Dynamic Screen Display condition, and 45.48\% (SD=5.08) in the Google Glass condition. Results of a repeated-measures ANOVA test showed no significant differences between the three conditions \( F_{2,58}=0.11, p=0.9 \).

Clicker Usage:

We also extracted the number of times that the presenter used the remote control to navigate through the topics. Average number of clicks for each slide was 9.09 (SD=0.94) in the Control condition,
0.38 (SD=0.18) in the Dynamic Screen Display condition, and 0.29 (SD=0.17) in the Google Glass condition. This shows that with speech-based tracking the presenter used the clicker very rarely.

5.2.8 Qualitative Findings

Our semi-structured interviews with the presenter participants were transcribed and coded using thematic analysis techniques [2]. From our analysis, we derived two main themes related to the effects of the speech-based dynamic note display approach and different display modalities in the presenters’ experience.

Dynamic vs. Static Note Display:

During the interviews, we asked participants to choose between the dynamic note display approach that automatically highlights probable next topics and the static display that allows manual topic navigation using the remote control. 19 participants preferred the dynamic display, while 10 participants chose the static display and 1 had no clear preference. Compared to the static display, the dynamic note method required less navigation effort: “It is like someone is doing it for me, so it’s easier than the clicker” [P22]. Having to operate the remote control for navigation while speaking could also break the presentation flow: “In manual, after every point I would look down and press the button, so I felt that it was breaking [the flow] in between” [P14]. This could cause the presenters to “lose eye contact with the audience” [P13] or fail to recall their points due to increased cognitive load: “[With the dynamic display] the prompts were coming immediately and they were flowing, so that was good... [With the clicker] when I was doing it manually, I almost forgot what I had to say” [P16].

Several participants commended the clear indication of covered and next topics in the dynamic note displays: “It was even a tick, so I was like ‘okay that portion is done,’ and I’m not continuously trying to go through the slide and being distracted by it” [P25]. Three participants specifically
indicated that this feature helped increase their confidence during delivery, because it gave them an assurance of their performance on content coverage: “In this case, we are quite sure we’ve covered everything and that’s why it is moving to the next step” [P15]. Finally, one participant felt that the speech-based tracking method encouraged her to speak more clearly during delivery: “I feel like it keeps a tab on how loud you are, because if you talk too softly it is not being to pick you up and it is not getting to your audience either” [P29].

On the other hand, most participants who preferred the static display stated that they appreciated the complete control offered by the manual navigation method: “I honestly preferred to just click through it. That way I can control the speed, and I can take time if I need to” [P10]. The static note display was also described as “the most reliable” [P2], “well-coordinated” [P30] and a “classic practice” [P30] that the presenters were most familiar with.

The most cited issue with the dynamic note display was the timing at which the display scrolls up to the next likely topic. Many participants reported instances that the display scrolled up to the next topic while they had not finished the current topic: “I liked using speech, but the only thing is that it was moving too fast. So before I could cover all the parts, it had already gone to the next point and I had no idea what I was supposed to speak” [P2]. To address this problem, future work is required to appropriately adjust both the threshold used for determining content coverage and the scrolling speed.

*Screen-based vs. Head-mounted Display:*

Participants reported mixed feedback on the display modalities. 16 participants expressed preferences for the screen-based dynamic note display, while 13 participants preferred the Google Glass and 1 participant had no clear preference. The most common reason for favoring the Google Glass was its potential benefits to maintain eye contact with the audience: “You can see the audience and the audience won’t see whether you are looking at them or you are looking at the
material. So that’s the way of getting interaction with the audience” [P5]. Several participants indicated that this capability helped increase their confidence: “It gave me a confidence boost. You know when I am presenting and I am constantly referring to my notes, it says that I’m not prepared enough or I’m too nervous to remember, that kind of image” [P2]. The novelty of the Google Glass could also lead to confidence boost: “Having a cool gadget on helped with my confidence too. It makes me feel more legit” [P14].

On the other hand, many participants reported that the Glass was distracting, and thus they “can’t focus on the audience”. The Glass was also reported to be uncomfortable to use during delivery: “When I was reading through the Glass, I have to constantly concentrate on the things that are appearing on the Glass, and that was creating a fuzziness and a dizziness and made me uncomfortable” [P9]. As the Glass required presenters to look upward to view the screen, participants were also concerned about the audience perception: “I would automatically tilt a little up to see what to say, so for the audience it might look weird with the speaker just looking up the whole time” [P12]. Despite the negative experiences with the Google Glass, many participants expressed the desire to continue using it in future presentations after they have sufficient practice with the device.

5.3 Conclusion

In this chapter, I described the design and evaluation of IntelliPrompter, an intelligent note display system that uses a speech-based presentation tracking framework to determine content coverage and automatically adjust the note interface for easy identification of next speaking points. We explored two modalities for dynamic note display: a conventional computer screen and Google Glass. Results of our evaluation showed that our speech-based dynamic note display approach achieved a high accuracy of 75.65%. The dynamic screen-based system was rated significantly higher in terms of usability compared to the static system, demonstrating the effectiveness of our
approach in improving presenters’ experiences. Despite the novelty and unique affordances of a head-mounted display, presenters reported mixed feedback on this modality, highlighting issues related to distraction, form factors and audience connection. Judges also rated the quality of presentations delivered with Google Glass to be significantly worse than those in the dynamic screen-based and static note systems. Based on these results, we can conclude that, when appropriate note display modality is used, presentation tracking could effectively be employed to support intelligent teleprompting that significantly improves the presenter’s experience during presentation delivery.

For future work, we aim to improve the usability of the note display interface by tailoring the automatic scrolling speed as well as the font size and color of the notes for each user. We also plan to improve the accuracy and evaluate our system with more diverse presentation topics and wider range of speakers.
CHAPTER 6

Quester System

In this chapter, I present a technology to support dynamic speaker-led presentations, by having a system track the presenter's speech and automatically index supporting slides in real time. This system, called Quester, matches a presenter's speech to slides based on slide contents and notes, and presents the top-ranked slides along with content keywords to the presenter so he or she only has to skim a handful of the most relevant slides and click on the one they want to show. We evaluated Quester in simulated post-presentation QA sessions, but it can be used to support any kind of speaker-led dynamically-structured presentation, such as a lecture on a subset of topics in a slide deck based on student or audience requests. The content of this chapter is published in [5].

The contributions include:

1. Analysis of recordings of existing QA sessions;
2. Development of the Quester system to automatically suggest and display presentation content based on speech;
3. Validation of Quester in a comparative study that demonstrates its potential to improve the presenter’s experience and performance during QA sessions, compared to standard slideware practice.

1 This study was conducted in collaboration with Ha Trinh, Harriet Fell, and Timothy Bickmore.
6.1 Question Answering Sessions

6.1.1 Guidelines

Anthony et al. [2] compared the advice given in self-help guide books regarding performance during QA sessions with their observations in actual conference talks. They argued that the quality of presentations can be improved with rehearsal, but presenters find it difficult to speak ad-lib during QA sessions. They observed the tendency of question askers to refer to something that was mentioned in the presentation before asking their question, and that presenters often repeat or rephrase questions before answering, as recommended in guidebooks.

6.1.2 Analysis of Question Answering Sessions

To better understand the quantitative characteristics of QA sessions, we analyzed video recordings of 20 conference presentations and 5 lectures. Conference presentation samples were chosen from 3 different venues: CHI Conference, National Education Alliance for Borderline Personality Disorder (NEAB-PD), and Fostering Entrepreneurship in the Creative Economy (FECE). Lectures were chosen from the Simons Institute interactive learning series about different topics in computer science. We annotated the duration of presentations, number of slides, length of QA sessions, and the number of questions asked. For lectures, we examined all questions asked during and after the presentation.

Based on our analysis, a typical 20-minute conference presentation had 32 slides on average, while a typical 60-minute lecture had 51. Navigating through such a large slide deck to search for a specific slide would potentially be difficult and time consuming.
Table 5. Statistical analysis of conference presentations and lectures

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CHI</td>
<td>10</td>
<td>14:24</td>
<td>33.1</td>
<td>3:30</td>
<td>3.2</td>
<td>9</td>
</tr>
<tr>
<td>NEAB-PD</td>
<td>5</td>
<td>20:36</td>
<td>28.2</td>
<td>5:37</td>
<td>4.4</td>
<td>0</td>
</tr>
<tr>
<td>FECE</td>
<td>5</td>
<td>19:24</td>
<td>15.4</td>
<td>7:57</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Lectures</td>
<td>5</td>
<td>60:20</td>
<td>51.8</td>
<td>-</td>
<td>12.8</td>
<td>8</td>
</tr>
</tbody>
</table>

In total, 145 questions were asked. We categorized the presenters’ strategies for responding to questions into two groups: 1) going back to a slide presented before (Refer); and 2) showing an extra slide (Extra) (Table 5).

On average, each QA interaction took 59 seconds, including the question, optionally repeating the question by speaker, optionally searching for and displaying a slide, and answering the question. For 10% of questions, presenters navigated to a slide (either previously presented or extra) to answer the questions. In 70% of these cases, the speakers moved though slides linearly. In the remaining 30%, they exited the presenter view to search for a slide. Navigating to a slide on average took 13.1 seconds, which comprises 22% of each QA interaction duration.

### 6.2 Design of Quester

Informed by our formative studies, we developed Quester, a question answering support system as an add-in for PowerPoint. Quester supports fast, nonlinear access to relevant presentation content by listening to speech input and dynamically suggesting the most related slides and note sections. Speech input can be received directly from the audience asking a question or from the presenter repeating or rephrasing the question. The system ranks presentation slides based on their textual content, slide notes, and metadata of slides’ visual elements. To further assist presenters in identifying target content, it also highlights the section of the slide notes that is most related to the
uttered question. Quester also provides an interface for presenters to privately preview and search for relevant content without exposing the search process to the audience.

6.2.1 Note Authoring Environment

Similar to previous studies we provided a note authoring environment. In addition to segmenting notes into topics, this also makes it possible to separate extra notes from the main presentation notes. Extra notes may contain supplemental material that are not presented during the main delivery, but the presenter can refer to them during QA. They are distinguished from main notes with a different font color (Figure 18).

6.2.2 Presentation Mode

During the presentation, Quester displays the slide, notes, and timing information, similar to PowerPoint’s Presenter View. It also provides a slide navigation pane, as shown in Figure 18.

Slide Navigation Pane:

Located at the bottom of the Presenter View, the slide navigation pane displays 5 slide thumbnails at each moment, but the presenter can scroll to see the rest of slides. The number of visible slide
thumbnails was chosen empirically based on trade-offs between quick accessibility and visual and cognitive load. Corresponding slide numbers are shown above each slide thumbnail. The slide navigation pane is hidden by default to avoid distraction during the main presentation, but it is shown when the presenter navigates to the last slide that he/she has planned to present. Users can set the index of their last slide in the authoring environment and use the rest of slides as extra slides.

The order of the slides displayed in the navigation pane is based on their semantic closeness to the speech input. It changes dynamically as new speech input is received and processed (with a 500 ms delay). To minimize potential distraction caused by rapid changes in the order of slide thumbnails, Quester does not change the slide orders once they appear in the visible portion of the Navigation pane. To indicate the degree of semantic match of each slide to recent speech, the system displays green bars with varying sizes above each slide thumbnail.

Presenters can hover over each slide thumbnail to privately see a preview of its content. This preview appears as a modal window over the presenter view and contains the slide and notes in

Figure 19. (Left) Slide Preview displayed when the presenter hovers the mouse over a slide thumbnail; (Right) Note pane is scrolled to display the most relevant note topic at the top with highlighted keywords
If the presenter wants to show a slide from the navigation pane to the audience, he/she can click on a thumbnail and the presentation will jump to that slide.

*Note Highlighting:*

When the speaker displays a slide using the navigation pane, the note topic most related to the speech content is detected and the keywords in that topic are highlighted. Highlighting is done by changing the color of keywords and increasing the font size. Also, the notes pane is scrolled to display that topic at the top (Figure 19).

In the next section, we describe our technical approach to estimating the semantic closeness between the slide content and speech content.

### 6.2.3 Semantic Match Scoring

We used a similar approach to previous studies to assign semantic scores. One difference is that content used for semantic matching includes slide text content, slide notes, and metadata about the visual elements in each slide. The metadata is generated by extracting the type of each element in the slide and adding corresponding descriptive words for the element to slide content. For example, if a slide contains a chart, the system adds the words “plot”, “graph”, and “chart” to the slide content. This metadata allows the system to handle questions that refer to specific visual elements of a slide (e.g. “*Can you go back to the slide with the graph on it?*”).

Also for this application, the system assigns slide-level and topic-level weights to each keyword based on its importance in the slide content and the note topic containing it, respectively.

### 6.2.4 Recency Model

The recency of keywords is important, so the keywords detected during a previous question do not affect the scores related to the current question being answered. To eliminate the effects of outdated keywords, the simplest solution is to require the presenter to manually reset slide scores before each
new question. However, this imposes an extra burden on the presenters and they might forget to do so. Therefore, we developed and evaluated two models that take into account the recency effect.

Our system stores detected keywords in a stack and if a keyword that is already put in the stack is detected again, it is moved to the top. Therefore, the most recent keywords will always be at the top of the stack. We examined 2 recency models:

Window model: Similar to previous work in word completion [116], we calculated slide scores based on only the top \( n \) keywords in the stack. For example, if \( n=2 \) then the model only considers the last 2 detected keywords.

Decay model: The model calculates slide scores after modifying the keyword weights based on the keyword positions in the stack. Keyword weights are modified using the following equation:

\[
    w = w_0 b^t \quad 0 < b \leq 1 \quad 0 \leq t
\]  

(7)

in which, \( w_0 \) is the original weight of keyword, \( b \) is the decay base, and \( t \) is the distance of keyword from the top of the stack. For \( b=1 \) the weights are the same as the original weights (i.e., no recency model). As \( b \) gets smaller, the decaying effect will be stronger and keywords detected further in the past will have a smaller effect on slide scores.

We performed theoretical evaluations of these two models using a corpus of presentations and questions collected from a user study that we describe in the next section.

6.2.5 Corpus of Presentations and Questions

To support both theoretical and user evaluations of different aspects of Quester, we developed a standardized corpus consisting of two short (7-slide) presentations on the topics of Exercise and Nutrition, along with a set of audience questions and extra QA slides for each.
Questions:
1- Can we eat a lot of fruit but not 2 liters of water?
2- You should consume 2 liters of water; does it change based on climate condition?
3- Should we have fluoride or chlorine in our water?
4- Is there something as too much water?

Slide Notes:

**Topic Title: How much water**
Water (on its own, or contained in food or other fluids) is essential part of daily diet; cannot be supplemented by anything else. Daily amount including water from food: 15 cups for average male; 11 for average female. Also depends on body size and climate. In rare cases, drinking an extreme amount in a short time can be dangerous. It can cause the level of salt, or sodium, in your blood to drop too low.

**Topic Title: Chlorine and Fluoride**
Cancer risk higher among drinkers of chlorinated water;
Fluoride in water:
- at low levels strengthens teeth
- at high levels can cause fluorosis, a discoloration of teeth
- at very high levels can cause skeletal damage

Figure 20. Sample Authored Slide and their Related Questions

To generate a corpus of audience questions for each presentation, we recruited 8 participants (college educated with presentation experience) to view videotaped presentations of a researcher giving each of the short presentations, then generating 10 questions each, resulting in an initial set of 80 questions per presentation. After removing redundant, unclear, and rhetorical questions, we further curated the questions, resulting in 40 audience questions per presentation. 4 of these 40 questions required display of slides previously presented in the main presentation (e.g. 'Can you
explain more about the graph on sleep?'), and 2 required display of extra slides (e.g. ‘Can you show me a sample weekly menu?’) (reflecting the Refer and Extra percentages in Table 5). The remaining questions could be answered using additional content, with or without displaying supporting slides.

To generate the extra / backup slides to provide support for the audience questions, we categorized the 36 questions for each presentation into 15 topics per presentation, and had two writers create an additional 15 slides with accompanying speaker notes for each presentation. The writers also created additional notes for the original 7 slides to help answer the 4 Refer questions for each presentation. Figure 20 displays a sample slide and the questions related to it.

The resulting slide notes for the Nutrition presentation contained 1562 words segmented into 55 topics. The slide notes for the Exercise presentation contained 1371 words in 58 topics. Our corpus roughly matched Table 1 on several metrics: the notes for each slide contained 73.3 words and 2.82 topics on average.

**6.2.6 Model Evaluation**

We used our presentation corpus to evaluate the accuracy of the matching algorithm using both recency models with different parameter values.

To ensure that the system performance does not depend on specific order of questions, we evaluated the system with 10 different randomized permutations of questions and calculated the average performance for all permutations. For each permutation of questions, we passed each question to the system as input to assign scores to the slides using different models and parameter values. The system then ranked the slides based on their scores.
**Measures:**

A common measure for evaluating the performance of QA systems is the mean reciprocal answer rank (MRR):

$$MRR_m = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{\text{Rank}_m(q)}$$  \hspace{1cm} (8)

In which, \( \text{Rank}_m(q) \) is the rank of the slide containing the answer to question \( q \) using model \( m \), and \( Q \) is the set of all questions.

We also measured the performance of our system using an accuracy measure more tailored to our application. We calculated the proportion of questions for which the slide containing the answer was included in the top \( p\% \) of slides with the highest match scores:

$$\text{accuracy}_{m,p} = \frac{|\{ \forall q \in Q | \text{Rank}_m(q) < p \}|}{|Q|}$$  \hspace{1cm} (9)

**Results:**

**Window Model:** We evaluated this model for different values of \( n \). We found that \( n = 4 \) resulted in the best MRR value of 0.70.

**Decay Model:** We evaluated this model for different values of \( b \). Based on the evaluation, \( b = 0.6 \) resulted in the best MRR value of 0.73. Figure 21 displays the MRR values of decay model for different base \( (b) \) values.

Using these optimized parameters (\( b=0.6 \) for decay model and \( n=4 \) for window model), we compared the accuracy of decay and window models against a system in which the scores are manually reset before each new question (Reset model). In our interface, we display the top 5 slides at any moment which corresponds to a \( p \) value of 22%. As shown in Figure 22, the accuracy of the Decay model for \( p=22\% \) is 90.2% compared to 87.1% for the Window model and 95% for the Reset model. The actual accuracy of the system is expected to be lower than these values due to
ASR errors. Based on these evaluations we decided to use the decay recency model with $b=0.6$ in our system.

Figure 21. Performance of the decay recency model with different base values

Figure 22. Comparison of different recency models
6.3 User Evaluation of Quester

To examine the effectiveness of dynamic speech-based slide matching in Quester in improving presenters’ experience and performance during a QA session, we conducted a user study comparing Quester against the static slide navigation supported in most conventional slidesware.

For our Control condition (the Static Navigation), we extended PowerPoint with a Navigation pane similar to the one in Quester, but which always displays slides in the same linear order that they appear in the slide deck. As in Quester, the presenter can hover the mouse over each slide thumbnail to preview the slide and notes. Once the presenter clicks on a slide thumbnail to show the slide to the audience, the system displays the corresponding notes without highlighting the most related note topic.

6.3.1 Procedure

We asked each participant to rehearse and deliver two 3-minute presentations on the topics of Exercise and Nutrition, using the PowerPoint slide decks and notes prepared in our QA corpus. After each presentation, presenters engaged in a brief QA session during which they were asked to answer a set of 8 questions randomly selected from our 40-question corpus for each presentation, using a different QA support system. For each 3-minute talk, the participants were asked to only present the first 7 slides in the slide deck. The remaining 15 slides were extra material that could be used during the QA session.

During the delivery and QA session, the experimentation room was set up with a 42-inch display monitor for slide projection, a tripod-mounted microphone for recording the presenter’s speech, and a secondary 15-inch computer screen displaying the Presenter View.
The study was a within-subject, single-session design in which each participant tested both conditions: *Quester vs. Static Navigation*. Each session lasted approximately 90 minutes. The ordering of the conditions and the slide decks were randomly assigned and counterbalanced.

At the beginning of the session, we introduced participants to the task of delivering two short presentations followed by QA sessions. For each condition, we gave participants 20 minutes to review the presentation content and rehearse their talk. We instructed them to review all 22 slides and notes, to prepare for both the presentation and the QA session. All participants confirmed that 20 minutes was sufficient for them to prepare.

Following each rehearsal, we gave participants brief instructions on how to use one of the two QA support systems and asked them to practice answering a sample question using the system. We asked participants to repeat the question aloud using their own words, before answering the question.

Following each practice round, we asked participants to give their final, videotaped presentation in front of the experimenter. After each presentation, the experimenter read aloud the randomly generated set of 8 questions, one at a time, for the presenter to answer. The presenter could ask the experimenter to repeat a question if they did not hear it clearly, and could skip any questions if they did not know the answer. Each QA session lasted approximately 5 minutes and was videotaped, with self-report measures gathered after each session. After the participant completed both conditions, we conducted a semi-structured interview, prompting for comparisons of the two QA support systems and suggestions for improvement.


6.3.2 Participants

We recruited 16 students and professionals (10 female, 6 male, ages 19-32, mean 23), with backgrounds in science, technology, business and arts. Of these participants, 2 were categorized as low competence public speakers, 5 were categorized as high competence public speakers, and 9 had moderate competence, according to the Self-Perceived Communication Competence Scale [69]. Participants were compensated for their participation.

6.3.3 Self-report Measures of Presenter Experience

We assessed the presenters’ experience with the QA support systems using the following self-report measures.

*Absolute Usability Rating of QA Support System:*

We asked presenters to rate the usability of each QA support system after each QA session, using a 7-item, 7-point composite scale, as shown in Figure 23.

![Figure 23. Presenters’ absolute usability ratings of two QA support systems (* indicates statistical significances)](image)
Figure 23 shows the results of the presenters’ absolute usability ratings of the two QA support systems. Results of a repeated-measures ANOVA test on the composite usability rating (Cronbach’s α=0.93) showed that Quester was rated significantly higher than the Static Navigation condition ($F_{1,15}=10.337$, $p=.006$, partial $\eta^2=.408$).

Results of Wilcoxon signed-rank tests on individual scale items showed that Quester was rated significantly better than the Static Navigation condition on helpfulness ($Z=-2.55$, $p=.011$), satisfaction ($Z=-2.81$, $p=.005$), desire to continue using the system in future presentations ($Z=-2.50$, $p=.013$), and how obvious it was to see what the answer is ($Z=-2.97$, $p=.003$). There were also trending effects of the condition on ease of use ($Z=-1.95$, $p=.051$) and reliability ($Z=-1.80$, $p=.072$), in favor of Quester. No significant difference was found between the two conditions in the rating of distraction ($p=.258$).

**Relative Usability Rating of QA Support Systems:**

After presenters completed both conditions, we asked them to directly compare and indicate which of the two QA support systems performed better on the same 7 criteria included in the Absolute Usability Rating.

Results of Chi-square tests revealed significant preferences of presenters towards Quester on all 7 criteria. Compared to the Static Navigation condition, Quester was rated to be significantly better in terms of ease of use ($\chi^2(1)=9.0$, $p=.003$), reliability ($\chi^2(1)=6.25$, $p=.012$), distraction ($\chi^2(1)=4.0$, $p=.046$), helpfulness ($\chi^2(1)=12.25$, $p<.001$), satisfaction ($\chi^2(1)=12.25$, $p<.001$), desire to continue using the system in future presentations ($\chi^2(1)=12.25$, $p<.001$), and how obvious it was to see what the answer is ($\chi^2(1)=12.25$, $p<.001$).

In summary, results of both absolute and relative usability ratings demonstrated a strong preference of the presenters for Quester.
Table 6. Presenters’ self-perceived ratings of their QA performances (Mean (SD) and p-value of Wilcoxon tests)

<table>
<thead>
<tr>
<th>Rating on scale 1-7 (1 - not at all, 7 - very much)</th>
<th>Static</th>
<th>Quester</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>How <strong>nervous</strong> were you…?</td>
<td>3.44 (1.55)</td>
<td>2.94 (1.34)</td>
<td>.143</td>
</tr>
<tr>
<td>How <strong>competent</strong> were you…?</td>
<td>4.63 (1.46)</td>
<td>5.25 (1.24)</td>
<td>.249</td>
</tr>
<tr>
<td>How would you rate the <strong>overall quality</strong>…?</td>
<td>5.06 (1.12)</td>
<td>5.31 (1.19)</td>
<td>.434</td>
</tr>
</tbody>
</table>

### 6.3.4 Self-perceived Rating of QA Performance

After each QA session, we asked presenters to assess their own QA performance using a 3-item, 7-point scale, evaluating their self-perception of their nervousness, competency, and overall quality of their performance.

As shown in Table 6, results of the ratings showed improvements in all measures with Quester. However, results of Wilcoxon signed rank tests showed no significant differences between the two conditions in any measures.

### 6.3.5 Objective Measures of Presenter Performance

We analyzed video recordings of all QA sessions and calculated the following objective measures of the presenter’s system usage and performance:

*Utilization Rate*: measured as the percentage of times that the presenter used the QA support system to answer a question, either by navigating through slides and notes to search for the correct answer and/or clicking on a slide to show to the audience. In both conditions, we asked participants to use the QA support systems in any way they wanted, and they could always choose to answer a question without using the systems.
Correctness of Answers: measured as the percentage of questions that were correctly answered, either with or without using the QA systems.

Efficiency: For each correct answer found, we measured the delay time (in seconds) taken from the moment the presenter finished repeating the question until the moment he/she started answering the question.

Number of Incorrect Slides Shown: measured as the number of incorrect slides shown to the audience before reaching a correct slide. This metric was calculated each time the presenter showed a correct slide to the audience.

A research assistant annotated the QA session recordings to check for the correctness of the presenter’s answers and the delay time for each correct answer. For each question, the annotator compared the presenter’s answer to the target answer included in the speaking notes. An answer was considered correct if it covered the main points in the target answer in sufficient details (but not necessarily word-for-word). 12.5% of the recordings (4 out of 32 recordings, 2 from each condition) were randomly chosen and annotated by another annotator to assess the reliability of our coding. The inter-rater reliability was adequate, with Cohen’s Kappa of 0.73 for correctness, and intraclass-correlation of 0.96 for efficiency.

We performed Shapiro-Wilk tests to check the normality of our data. Our data for all measures were not normally distributed, thus we analyzed our data using non-parametric tests (Wilcoxon signed rank tests). Results of the objective measures are shown in Table 7. There was a significant difference in utilization rate between the two conditions (Z=-1.97, p=.048), in favor of Quester. There was also a significant difference in the answer response delay time (for correct answers) between the two conditions (Z=-2.04, p=.041), in favor of Quester. There were no significant differences between the conditions on the Correctness of Answers (Z=-0.77, p=.441), or the Number of Incorrect Slides Shown (Z=-1.35, p=.176).
Table 7. Objective measures of presenters’ system usage and performance during QA sessions (Mean (SD) and p-value of Wilcoxon signed rank tests)

<table>
<thead>
<tr>
<th>Measures</th>
<th>Static</th>
<th>Quester</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilization Rate (%)</td>
<td>59.37 (36.37)</td>
<td>74.22 (32.1)</td>
<td>.048</td>
</tr>
<tr>
<td>Correctness of Answers (%)</td>
<td>77.53 (23.0)</td>
<td>81.11 (16.35)</td>
<td>.441</td>
</tr>
<tr>
<td>Efficiency (seconds)</td>
<td>16.28 (11.15)</td>
<td>8.36 (6.43)</td>
<td>.041</td>
</tr>
<tr>
<td>Number of Incorrect Slides Shown</td>
<td>2.09 (3.25)</td>
<td>0.46 (0.69)</td>
<td>.176</td>
</tr>
</tbody>
</table>

System Accuracy:

To evaluate the performance of Quester, we analyzed system logs and calculated the system accuracy at both slide and note topic levels. *Slide-level accuracy* was measured as the percentage of times that the correct slide appears in the top 5 recommended slides for each question. *Topic-level accuracy* was measured as the percentage of times that the correct note topic was highlighted when the presenter clicked on the correct slide for a question. The average *slide-level accuracy* was 84.37% (SD=17.38), and the average *topic-level accuracy* was 74.05% (SD=32.31).

Results of Pearson’s correlation test showed significant correlation between the *accuracy* and the *composite usability rating* of the Quester system (r = .565, n = 16, p = .023). There was also a significant correlation between the *accuracy* of the system and the *correctness of answers* given by the presenters (r=.754, n=16, p=0.002) but there were no significant correlations between the *accuracy* and other objective measures.

6.3.6 Qualitative Findings

We performed high-level coding of our semi-structured exit interview transcripts and derived three main themes.
Impact on User Experience:

15 of 16 participants expressed strong preferences for Quester, citing its positive effects in time saving that encouraged them to actually use the system more: “The first system [Static], I didn’t use it much because you need to actually go in and search yourself. The other system [Quester] will do it for you and you will save time” [P1]. This is the most important benefit of Quester, as long delay time could “interrupt your train of thought” [P12] and lead to negative impressions from the audience. Quester was also found to help reduce the memorization load required to “remember where everything you put in the presentation is” [P12]. In addition, two participants specifically reported that our system helped increase their confidence: “When I had to look into the slides, there is a lot less lag time, and that lack of lag time made me feel more confident, because I wasn’t keeping the audience stalling” [P11]. One participant, however, preferred the Static Navigation system over Quester, because he felt that it was more “organized” [P2] and easy to simply skip through slides in order.

Level of Distraction:

11 of 16 participants indicated that they experienced no distraction with Quester, as one participant explained: “I wasn’t looking at the computer until I had repeated the question, at which point it was already adjusted to include the right proportion of the slides and the likelihood. So it wasn’t distracting to me” [P11]. Several participants specifically reported that they felt the static navigation system was actually more distracting, because: “While scrolling down I happened to see all the slides. And my mind started thinking about all the slides” [P6].

However, 2 participants reported that the changing green bars that indicate slide likelihood was confusing and distracting. In addition, 3 participants mentioned that when the system failed to show correct slides, it was difficult and distracting for them to search for relevant slides themselves, as the slides “were not in the orderly manner” [P5]. One potential solution is to combine both
dynamic and static navigation approaches, allowing the presenter to switch back to static navigation when the dynamic slide recommendations are not accurate.

*Suggestions for Improvement:*

Some participants had suggestions for improving the appearance of slide recommendations. They proposed using color codes rather than varying sizes of the green bars to indicate different ranges of slide probability. 2 participants also suggested simply displaying the recommended slides in the descending order of probability. Participants also proposed novel ways to extend the system functionality. For example, the system could automatically display relevant information from external resources, such as webpages or existing documents.

**6.4 Conclusion**

In this chapter, I presented Quester, a system that provides dynamic speech-based slide ranking for speakers giving non-linear presentations, such as during post-presentation QA sessions. In our evaluation study, presenters rated Quester significantly higher, used it significantly more, and answered audience questions significantly faster, compared to a system that only provided linear slide navigation. By utilizing the presentation tracking, Quester improved both the presenter’s experience and performance during QA sessions.

Support for non-linear presentation formats represents an important area for future research, given that current slideware—which enforces staid, linear presentation formats—has not fundamentally changed in several decades. Future directions of research include novel forms of displays and prompts to help presenters give adaptive but coherent presentations, with supporting media (e.g. external resources or dynamically-updated data visualizations) automatically provided by the system.
CHAPTER 7

RoboCOP System

The lack of feedback on the content of presentations is one of the shortcomings of currently available rehearsal support systems. In this chapter, I demonstrate the effectiveness of presentation tracking in providing content-based feedback during presentation rehearsal. The tracking framework is integrated into RoboCOP, a rehearsal environment in which presenters rehearse their talks in front of an automated speaking coach, which acts as both an audience and an empathetic expert coach that provides spoken feedback on multiple facets of presentation delivery. To provide the greatest audience realism through “sense of presence”, and to give the speaker a focal point for his or her rehearsal without the clumsiness of a head-mounted display, we use an anthropomorphic robotic head as an embodiment for the rehearsal coach. The coach can identify a range of speaker

Figure 24. Presenter rehearsing her talk in front of the robotic coach
behaviors automatically, including speech quality, content coverage, and head orientation behavior, and provides feedback in a natural conversational manner.

In the rest of this chapter, I present an exploratory study of human rehearsal coaches that informed the design of RoboCOP system, describe the design and implementation of this system, and report on an evaluation study in which presenters rehearsed with and without the automated robotic coach and had their resulting presentations rated by a panel of human judges. The content of this chapter is published in [110].

7.1 Understanding Rehearsal Coaching Practice

To motivate and inform the design of our rehearsal coaching system, we conducted an exploratory study to understand the practices of presentation coaching. Our aim was to identify the categories, structure, language and frequency of feedback offered by human coaches during presentation rehearsal.

7.1.1 Participants

We recruited 8 professors (5 male, 3 female) from the health science, computer science, music and theatre departments at our university. All participants were experienced in either teaching public speaking classes or mentoring students on their oral presentations.

7.1.2 Procedure

Each study session lasted approximately 1 hour, in which participants were asked to give coaching feedback during presentation rehearsals. Each rehearsal was 7-10 minutes long on general knowledge topics (France, Italy, Lions and Tigers), and was either pre-recorded or a live rehearsal.

____________________

1 This study was conducted in collaboration with Ha Trinh, Darren Edge, and Timothy Bickmore.
The recorded rehearsals were randomly selected from a pool of 24 videotaped rehearsals of 12 students and professionals with varying levels of presentation experience, collected from our prior studies on presentation technologies. In those studies, participants were given 30-60 minutes to review pre-made slides and notes before delivering their talk in front of a camera. The live rehearsals were given by research assistants in our lab, who were given the presentation materials five days in advance and were instructed to prepare in any ways they wanted before practicing their talk with a coach. To mitigate learning effects, each assistant only performed two rehearsals.

In each study session, we asked coach participants to watch two different rehearsals from different speakers, who were unknown to them. The first rehearsal was an early-stage rehearsal, during which participants were asked to give preliminary feedback on how to improve the speaker’s performance. They were instructed to interrupt the speaker at any time during the rehearsal, and give any verbal feedback that they thought would be useful. The second rehearsal was a complete dry-run, for which we asked participants to wait until the end of the presentation and give all their summative feedback for the entire talk.

Prior to each rehearsal, we explained the goal and the target audience of the presentation to participants, and gave them a handout of the slides and notes specifying the key points that should be covered in the talk. In the case of videotaped rehearsals, we asked participants to imagine that the speaker was present in the room and to speak their feedback directly to the speaker.

7.1.3 Findings

We recorded and transcribed all participants’ coaching feedback, resulting in a total of 78 early-stage feedback samples and 8 dry-run feedback samples. During early-stage rehearsals, each coach gave an average of 1.6 feedback samples per slide (SD = 1.2). Most of the early-stage feedback occurred at the end of a slide. Coaches often gave highly detailed feedback (mean length of early-
stage feedback = 103.7 words, mean length of dryrun feedback = 385.7 words), which comprised
descriptions of the speaker’s performance, actionable suggestions with explanatory justification,
and positive reinforcement. Feedback messages were often structured using the “feedback
sandwich” technique [26], starting with positive messages before proceeding to suggestions for
improvement.

To identify feedback categories, we annotated each feedback sample with a category code. Table
8 shows the 11 common feedback categories grouped into four main themes, along with their
frequencies of occurrence in our feedback corpus. The experts provided feedback on a wide range
of topics, spanning talk planning, organization and slide design (44.6% of comments), content
coverage (4.1%), body language and eye contact (19.8%), and speech quality (31.5%), with the last
category further broken down into language and pronunciation (11.6%), speaking rate (8.3%), use
of “fillers”, such as “umms” and “ahs” (7.5%), and voice pitch variety (4.1%). When the specific
phrasing of expert feedback was particularly clear and helpful, we noted these phrases as candidates
for inclusion in an automated rehearsal coach.

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talk Planning, Organization and Design</td>
<td>44.6</td>
</tr>
<tr>
<td>Goal &amp; Audience Benefits</td>
<td>19.9</td>
</tr>
<tr>
<td>Organization</td>
<td>6.6</td>
</tr>
<tr>
<td>Introduction &amp; Close</td>
<td>8.3</td>
</tr>
<tr>
<td>Slide Design</td>
<td>9.9</td>
</tr>
<tr>
<td>Speech Quality</td>
<td>31.5</td>
</tr>
<tr>
<td>Language / Pronunciation</td>
<td>11.6</td>
</tr>
<tr>
<td>Speaking Rate</td>
<td>8.3</td>
</tr>
<tr>
<td>Filler Rate</td>
<td>7.5</td>
</tr>
<tr>
<td>Pitch Variety</td>
<td>4.1</td>
</tr>
<tr>
<td>Nonverbal Behavior</td>
<td>19.8</td>
</tr>
<tr>
<td>Body Language</td>
<td>9.9</td>
</tr>
<tr>
<td>Eye Contact</td>
<td>9.9</td>
</tr>
<tr>
<td>Content Coverage</td>
<td>4.1</td>
</tr>
</tbody>
</table>
7.2 Design of RoboCOP

Informed by findings from our exploratory study, we developed RoboCOP (Robotic Coach for Oral Presentations), an automated anthropomorphic robot head for presentation rehearsal. The robot plays the role of a coach who actively listens to the presenter’s spoken rehearsals and offers detailed spoken feedback on five key aspects of presentations: content coverage, speaking rate, filler rate, pitch variety, and audience orientation (which is considered a proxy for eye contact). These metrics were chosen based on our exploratory study and previous automatic presentation quality assessment studies [8, 44]. In addition, the coach also provides high-level advice on the presentation goal and audience benefits, as well as talk organization, introduction and close. Our aim was to simulate the interactive nature and feedback mechanisms of rehearsing in front of a live audience, while mitigating public speaking anxiety that often arises when performing with actual human audiences. Unlike existing virtual audience-based rehearsal systems that provide indirect feedback through non-verbal behavior, our robot provides detailed, structured, actionable and empathetic feedback that resembles the behavior of human coaches. I first present an overview of the presentation preparation process with RoboCOP, followed by descriptions of its core components.

Figure 25. RoboCOP Microsoft PowerPoint add-in with (a) note authoring pane and (b) presenter view
7.2.1 Presentation Preparation with RoboCOP

Prior to spoken rehearsals, RoboCOP enables the presenter to prepare speaking notes for each slide using our topic-based note authoring interface (Figure 25). Implemented as an add-in for Microsoft PowerPoint 2016, our note authoring tool allows the presenter to segment the speaking notes of each slide into a series of key topics. The presenter can enter a short title for each topic, along with detailed notes specifying what they intend to say about it. During rehearsal, our system tracks the presenter’s speech to determine which topics have been covered on each slide, and provides feedback on content coverage accordingly.

Once the presenter is ready for the first spoken rehearsal, he/she activates the Slide Walkthrough mode by clicking on the corresponding control in the PowerPoint ribbon. In this mode, the presenter practices verbalizing slides while receiving preliminary feedback from the robotic coach at the end of each slide. At the beginning of this mode, the coach engages presenters in a short introductory dialogue before proceeding to the rehearsal. This simple dialogue serves three purposes: (1) establishing the role of the robot; (2) familiarizing the presenters with the concept of talking to and receiving feedback from the robot; (3) prompting them to keep in mind the overarching goal of their presentation and their target audience while presenting. In this interaction, user input is limited to acknowledgment utterances that only serve to advance the dialogue.

To facilitate the rehearsal, we provide the Presenter View, which displays all topic notes of a slide on a single note page, along with the slide and timing information (Figure 25). During the rehearsal, the coach acts as an attentive audience and offers verbal feedback at the end of each slide on five key presentation quality metrics: content coverage, speaking rate, filler rate, pitch variety and audience orientation.

Once the presenters master each individual slide with the Slide Walkthrough mode, they can proceed to the Dry Run mode to perform a complete practice talk from beginning to end, without
interruption from the coach. At the beginning of this mode, the coach also engages the presenters in an introductory dialogue, reminding them to pay special attention to verbal transitions between slides and the presentation timing. The coach also encourages the presenter to prepare for a strong introduction and close. During the rehearsal, the coach actively listens to the presenter’s speech, but does not give feedback at the end of each slide to avoid interrupting the presentation flow. Instead, she provides summative feedback on the overall presentation at the end of the talk, focusing on the same five categories as in the Slide Walkthrough mode.

We now describe the three core components of our system, including the robotic coach, the presentation quality assessment module, and the feedback generation module.

### 7.2.2 The Robotic Coach

We used Furhat [1], a human-like robot head, as our presentation coach. Furhat consists of an animated face model that is back-projected onto a 3D translucent mask (Figure 24). Our coach speaks using a female synthetic voice from CereProc [6], with synchronized lip movements. She is capable of displaying a variety of non-verbal behaviors while speaking, including facial expressions of affect (smile, neutral, concern), eyebrow movements, directional gazes and head nods. Most of her non-verbal behaviors are automatically generated using the BEAT text-to-embodied speech system [21]. Human-robot conversations are scripted using our custom scripting language based on hierarchical transition networks. Users contribute to the conversation via speech input. However, the current system does not incorporate natural language understanding functionality. Thus, the coach does not attempt to interpret the user’s responses, and simply relies on speech pauses to advance the dialogue.

While interacting with the presenter, the coach exhibits two types of listening behavior. First, the system uses a Microsoft Kinect 2 camera to track the location and rotation of the presenter’s head. As the presenter walks around during the presentation, the robot head moves so as to maintain its
eye gaze in the presenter’s direction. Second, the robot provides non-verbal backchannel feedback in the form of head nods at appropriate times based on acoustic features of the presenter’s speech. Using a similar approach to [65], we detect two prosodic cues, including raised loudness and lowered pitch. To identify these events, we continuously process the last 2 seconds of speech at every 500-millisecond interval. We track prosodic events occurring at least 500 milliseconds before the end of the speech sample. If the average intensity during the last 100 milliseconds of the voiced part of the sample is higher than the 99th percentile of the intensity contour, we signal a raise in loudness. If the average pitch value for the same period is lower than the 23rd percentile of the pitch contour, we signal a lowered pitch.

### 7.2.3 Presentation Quality Analysis

During the presentation, the system calculates the pitch range, speaking rate, filler rate, and audience orientation every 20 seconds and reports the average of these values at the end of each slide. It also reports the content coverage for each slide; determining whether the key ideas in the slide notes were spoken by the presenter.

**Content Coverage:**

To measure the content coverage for each slide, we used my presentation tracking framework.

**Speech Quality Features:**

To determine the speaker’s ranges for pitch and voice intensity, we calibrated the system at the beginning of each session. We asked the speaker to read two short prepared lines of text and record their voice. We used these recordings to extract the pitch and intensity contours in Praat [13]. This information was used for setting the silence threshold, which is 25% of the difference between the 1st and 99th percentile of intensity, and the thresholds used for identifying the prosodic cues for listening behavior.
Pitch is estimated using an autocorrelation method with a floor value of 75 Hz and a ceiling value of 500 Hz, which are Praat’s default settings. While the speaker’s pitch may vary based on speech content, previous studies show that the overall pitch variety is significantly correlated with speech quality [8]. To measure *pitch variety*, we calculate the difference between the 90th and 10th percentile (80% range) and the 95th and 5th percentile (90% range) of pitch in Hertz and semitones. Semitone is a logarithmic scale which shows the perceived pitch variation, and it can remove cross-gender differences [85]. We also calculate the Pitch Dynamism Quotient (PDQ) by dividing the standard deviation of pitch by pitch mean values. PDQ has been used as a normalized measure for pitch variation [85].

We used the method in [25] to estimate the *speaking rate*. Segments of speech with intensity values lower than the silence threshold or undefined pitch values are marked as unvoiced segments. Peaks in the intensity envelope of the voiced parts of the signal are identified and those that are at least 2 dB higher than their succeeding peaks are extracted as syllable nuclei. To calculate the speaking rate, we divide the number of syllables by the speaking time. Speaking time is defined as the total audio sample length minus the sum of length of all pause segments. Pause segments longer than one second are considered as one second to remove the effect of long pauses on speaking rate.

To measure the *filler rate*, we use the IBM Watson ASR to transcribe the speech and count filled pauses, such as “um” and “uh”, and the word “like” in the speech transcription. Although “like” can be used as a non-filler word, it is the most commonly used filler word [53]. Previous research shows that simply counting all occurrences of potential filler words can result in approximately 70% accuracy in filler rate measurement. Using language processing rules to filter non-filler usages can only reduce the error rate from 30% to 19%, at the expense of much more complex algorithms [49]. The total number of fillers is divided by the speaking time in minutes to determine the filler rate in fillers/minute.
**Audience Orientation:**

As a proxy for eye contact measurement, our system uses Microsoft Kinect to track the speaker’s head orientation to determine whether their focus of attention is on the robot audience rather than on the projected slides or speaking notes. The Kinect is located behind and above the robot. Previous research [55] has shown that using head pose could yield acceptable accuracy for real-time estimation of attentional focus, without the expense of bulky eye trackers. The audience orientation ratio is calculated as the amount of time that the speaker is looking at the robot while speaking divided by total speaking time.

**Discretizing the Quality Measures:**

In order to provide feedback on quality measures, we needed to set proper thresholds and ranges. Similar to [93, 107], we defined these values using empirical data. We conducted a small user study in which we asked 8 participants to rate the speech quality of presentation samples randomly selected from a corpus. The corpus includes 696 samples, each 20 seconds long, which were extracted from 30 presentation recordings of 21 different speakers. We automatically extracted the speech quality measures for these samples. The samples were ordered based on the values of speech quality features and grouped into 20 bins. Each participant watched 20 samples, one randomly selected sample from each bin, and rated the speaking rate, pitch variety, and usage of fillers. We also recruited an additional 8 participants to rate the presenter’s eye contact in 20 presentation recordings.

We grouped the values of speech quality measures from samples based on participants’ ratings. ANOVA tests showed significant differences among group means for speaking rate, filler rate, and 90% pitch range in Hertz. The results showed no significance for 90% and 80% pitch range in semitones, 80% pitch range in Hertz, and PDQ. Therefore, we used 90% pitch range in Hertz as the pitch variety measure. Based on the participant ratings, we set ranges and thresholds for each
presentation quality measure, as shown in Table 9. We evaluated the performance of our classifications by comparing the results of our automatic classifications against participants’ ratings. Results of our evaluation showed that the system achieved 58.6% F1 for filler rate, 65.1% F1 for pitch variety, 46.1% F1 for speaking rate, and 84.3% F1 for audience orientation.

**Table 9. Ranges and thresholds for speech quality metrics**

<table>
<thead>
<tr>
<th>Measures</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking Rate (syl/s)</td>
<td>[0, 3]: slow</td>
</tr>
<tr>
<td></td>
<td>(3,5): good</td>
</tr>
<tr>
<td></td>
<td>[5, ∞): fast</td>
</tr>
<tr>
<td>Fillers (fillers/minute)</td>
<td>[0,5): good</td>
</tr>
<tr>
<td></td>
<td>[5,15): some</td>
</tr>
<tr>
<td></td>
<td>[15, ∞): many</td>
</tr>
<tr>
<td>Pitch Variety (Hz)</td>
<td>[0, 120): monotone</td>
</tr>
<tr>
<td></td>
<td>[120, ∞): good</td>
</tr>
<tr>
<td>Audience Orientation</td>
<td>[0, 0.4): low</td>
</tr>
<tr>
<td></td>
<td>[0.4,1]: good</td>
</tr>
</tbody>
</table>

### 7.2.4 Identifying Performance Trends

We also determine the trends for each speech quality measure at both the slide level and overall presentation level, which can be used to generate feedback on performance trends, as described in the next section. For overall presentation level, we defined five different types of trend:

1. Flat Good: If the measure value in more than 80% of the slides is in the “good” range
2. Flat Bad: If the measure value in more than 80% of the slides is not in the “good” range
3. Improving: If the measure value in the first 40%-60% of the slides is not in the “good” range but in the rest of slides is in the “good” range
4. Degrading: If the measure value in the first 40%-60% of the slides is in the “good” range but in the rest of the slides is not in the “good” range
5. Variable: Other cases
For slide-level trends, we compared the performance of two consecutive slides and defined five trend types:

1. Significant Improvement: If there is a change in the range of the measure, in the positive direction
2. Slight Improvement: If there is no change in the measure range, but there is at least 10% improvement in the measure value
3. Flat: If there is no change in the range of the performance
4. Slight Degradation: If there is no change in the measure range, but there is at least 10% degradation in the measure value
5. Significant Degradation: If there is a change in the range of the measure, in the negative direction

7.2.5 Feedback Generation

Using the output from the presentation quality analysis, we automatically generate two types of verbal feedback, including slide-level feedback provided at the end of each slide in the Slide Walkthrough mode, and presentation-level feedback provided at the end of the talk in the Dry Run mode. Our aim was to offer constructive coaching feedback that combines both contextualized suggestions for improvements and positive reinforcement to build speaker confidence. Our feedback generation module, described next, is developed based on the standard Natural Language Generation (NLG) pipeline [89].

7.3 Comparison of Feedback Modalities

To investigate the effects of RoboCOP on the presenter’ experience, we conducted a user study comparing our robot-based coaching feedback against visual feedback (Graphic condition, Figure 26) and verbal feedback without robot (Voice Only condition). Our aim was to evaluate the effects
of both the physical embodiment of the rehearsal coach and the use of verbal feedback on the overall rehearsal experience of presenters (Figure 27).

Figure 26. Example of graphical feedback displayed at the end of each slide

Quantitative and qualitative results demonstrated the positive effects of the robot-based coaching approach to improve the overall rehearsal experience of presenters, compared to both the Voice Only and Graphic conditions. On the other hand, there were no major differences between the Voice Only and the Graphic conditions, except that participants were more engaged with the verbal feedback than with the graphical feedback.

7.4 Evaluation of RoboCOP

We conducted a user study comparing rehearsals with RoboCOP against rehearsing alone without coaching feedback (control condition). While our feedback modality study focused on the user experience of presenters, our aim in this study was to examine the effectiveness of the robotic coach in improving both the presenter’s experience and the presentation quality as perceived by an audience, when compared to existing rehearsal practices. We considered this as a significant step
towards validating the effectiveness of our system, and providing empirical evidence that automated feedback during presentation training can actually lead to increased presentation quality.

7.4.1 Procedure

We asked each participant to rehearse and deliver two 7-minute presentations on comparable topics (French and Italian Culture) in English using prepared PowerPoint slide decks and notes. Each slide deck contained 6 slides and approximately 600-word supporting notes, covering 17 key points. In one of the presentations, presenters were asked to rehearse with the robotic coach, while in the other presentation they rehearsed alone in front of a camera.

The study was a within-subject, counterbalanced design across two sessions. Each session lasted between 60-90 minutes, with 1 to 5 days between sessions. The ordering of the conditions (RoboCOP vs. Control) and the slide decks were randomly assigned and counterbalanced. The rehearsal and the final presentation were videotaped for later evaluation.

RoboCOP Session:

At the beginning of the session, we introduced participants to the common scenario of presenting using a pre-made slide deck, as well as the presentation goal and target audience. We instructed
them to cover all the key points in the notes, but not necessarily word-for-word. Following this introduction, we allowed them 15 minutes to review the slides and notes in PowerPoint, before performing two rounds of spoken rehearsal. In the first rehearsal, the participants used the Slide Walkthrough mode to practice presenting each slide and receiving the coach’s feedback at the end of each slide. In the second rehearsal, they were asked to perform a complete practice talk using the Dry-run mode and receive summative feedback from the coach at the end of their rehearsal. The entire rehearsal session lasted approximately 30 minutes and was videotaped. The experimenter was not present during the rehearsal. Following the rehearsal, participants were asked to deliver their final, videotaped presentation in front of the experimenter. The robotic coach was not present during the final talk. The session concluded with a semi-structured interview, eliciting the presenter’s experience of rehearsing with the coach and suggestions for improvement.

Control Session:

In this session, we asked participants to rehearse for their presentation without the presence of the robotic coach. We gave participants the same scenario as in the RoboCOP session, before giving them 15 minutes to review the slides and notes. The participants were then asked to perform two rounds of videotaped, spoken rehearsals that lasted approximately 30 minutes, before giving a final, videotaped presentation. In the first rehearsal, they were instructed to go through and practice each slide aloud. In the second rehearsal, they were asked to perform a complete practice talk from beginning to end, as if they were in front of their audience. The experimenter was not present in the rehearsal. Following the rehearsal, participants were asked to deliver their final, videotaped presentation in front of the experimenter. We concluded the session with a semi-structured interview.
7.4.2 Presenter Participants

We recruited 12 students and professionals with technical backgrounds and varying levels of presentation experience (3 female, 9 male, ages 22-28, mean 24). Of these, 7 were categorized as high competence public speakers and 5 had moderate competence according to the Self-Perceived Communication Competence Scale [21]. None of the participants interacted with the robotic coach prior to the study.

7.4.3 Presenter Measures

Presenters and their attitudes were assessed using the following self-report measures:

*State Anxiety:* Assessed prior to each presentation using the State Anxiety questionnaire [31].

*Speaker Confidence:* Assessed at intake and after each presentation using the Personal Report of Confidence as a Speaker questionnaire [83].

*Coach Rating:* Assessed after the final presentation in the RoboCOP session using a 6-item, 7-point scale questionnaire, as shown in Table 10.

7.4.4 Presenter Quantitative Results

Results of repeated measures ANOVA tests showed no significant effects of condition on state anxiety (F₁,₁₀=1.15, p=.31) or speaker confidence (F₁,₁₀=.25, p=.63).

Results of coach ratings showed that presenters were highly satisfied with the coach (M = 5.92, SD = 1.31), found her to be helpful (M = 5.83, SD = 1.4), likable (M = 5.67, SD = 1.87), and expressed high desire to continue working with her in their future presentations (M = 5.92, SD = 2.11). The rating with the lowest result was trustworthiness (M = 5.17, SD = 1.70), due to the inaccuracy of the coach’s feedback in some instances. These inaccuracies were found mainly in the audience.
orientation feedback and occasionally in content coverage feedback, resulting from the use of head orientation as a proxy for eye contact measurement and the imperfect automatic speech recognition.

### Table 10. Average Ratings of the Robotic Coach

<table>
<thead>
<tr>
<th>Rating of the Coach:</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Scale Measures from 1-7)</td>
<td></td>
</tr>
<tr>
<td>1 – not at all</td>
<td>7 – very much</td>
</tr>
<tr>
<td>How satisfied are you with the coach?</td>
<td>5.92 (1.31)</td>
</tr>
<tr>
<td>How much do you like the coach?</td>
<td>5.67 (1.87)</td>
</tr>
<tr>
<td>How much do you feel you trust the coach?</td>
<td>5.17 (1.70)</td>
</tr>
<tr>
<td>How helpful was the coach?</td>
<td>5.83 (1.40)</td>
</tr>
<tr>
<td>How much would you like to prepare future presentations with the coach?</td>
<td>5.92 (2.11)</td>
</tr>
<tr>
<td>How easy was it to use the coach?</td>
<td>5.25 (2.05)</td>
</tr>
</tbody>
</table>

#### 7.4.5 Evaluation of Presentation Quality

To evaluate the relative quality of the 12 pairs of videotaped presentations (RoboCOP vs. Control) that were delivered by our presenter participants, we recruited 12 judges (6 female, 6 male, ages 23-55, mean 31). Judges were students, researchers and professors with varying levels of presentation experience.

We asked each judge participant to watch two pairs of videotaped presentations and complete the following questionnaires:

*Absolute Rating of Presentation Quality*: Assessed after watching each presentation, using a 7-item, 7-point scale questionnaire evaluating engagingness, understandability, novelty, excitement, entertainingness, overall quality and desire to continue seeing similar presentations, as shown in Figure 28.
Figure 28. Absolute ratings of presentation quality for the Robot vs. Control conditions (* indicates significant differences)

Figure 29. Audience perception of presenters for the Robot vs. Control conditions (* indicates significant differences)
Audience Perception of Presenters: Assessed after watching each presentation, using a 7-item, 7-point scale questionnaire evaluating the presenter’s competency, engagingness, nervousness, understandability, excitement, entertainingness and overall satisfaction, as shown in Figure 29.

Relative Rating of Presentations: Assessed after watching each presentation pair from the same presenter (RoboCOP vs. Control), comparing the relative quality of each pair on six criteria adopted from [112], including: organization, content coverage, note reliance, speech quality, timing and pacing, and overall quality, as shown in Figure 30. Each criterion was judged on a 4-point ordinal scale of “no difference”, “slight difference”, “moderate difference”, and “substantial difference”, with an indication of the superior presentation, if any.

Each judge session lasted approximately 40 minutes. The ordering of the presentations was randomly assigned and counterbalanced across the judge participants.

Judge Rating Results

We performed non-parametric tests (Wilcoxon signed-rank tests) to examine the effects of our RoboCOP system on the judges’ ratings of presentation quality and presenters. Results are as follows:

Absolute Rating of Presentations: Figure 28 shows the judges’ absolute ratings of presentation quality for the two conditions. Judges rated presentations prepared with RoboCOP to be significantly more engaging (Z = -2.17, p = .03), novel (Z = -2.22, p = .027) and exciting (Z = -2.10, p = .036). There were no significant differences between the two conditions for the other four measures.

Audience Perception of Presenters: Figure 29 shows the judges’ ratings of the presenters for the two conditions. Judges rated presenters to be significantly more competent (Z = -2.34, p = .019),
and were significantly more satisfied with the presenters ($Z = -2.20$, $p = .028$) in the RoboCOP condition. No significant differences were found for other measures.

*Relative Rating of Presentations:* Figure 30 shows the judges’ relative ratings of presentation quality. There were significant differences on the judges’ ratings of *speech quality* ($p = .037$) and *overall presentation quality* ($p = .042$), in favor of the RoboCOP condition. No significant differences were found for organization, content coverage, note reliance, timing and pacing.

![Figure 30. Relative ratings of presentation quality for the Robot vs. Control conditions (positive values indicate the Robot condition is better, * indicates significant differences)](image)

*Content Coverage:* In addition to the judge ratings, two researchers in our team also independently annotated the final presentation recordings to check for content coverage. Each annotator was given a checklist of the 17 key points that should be covered in each presentation, and was instructed to award one point for each piece of content presented in sufficient details. We calculated the average content coverage score between the two annotators for each presentation and used them for comparison. The annotators had relative agreement of 89.7%.

109
Average content coverage was 91.9% (SD=4.18) in the RoboCOP condition, and 88.5% (SD=11.25) in the Control condition. Results of a repeated-measures ANOVA test showed no significant differences between the two conditions ($F_{1,11}=1.8$, $p=.21$).

To calculate the performance of the tracking system, we compared the automatic content coverage tags with human annotations. The average accuracy of the tracking system was 82.62% (SD=13.02).

### 7.4.6 Qualitative Findings

Semi-structured interviews with the presenters were transcribed, coded and categorized into three main themes relating to the effects of RoboCOP on spoken rehearsal experience, presentation quality, and speaker confidence.

*Facilitating Spoken Rehearsal with an Attentive Audience:*

Most participants appreciated the benefits of the robot’s presence in creating a more natural, interactive and motivating rehearsal environment: “She was a true virtual audience. She was attentive…which makes you feel like talking even more. You don’t feel like you are talking to air or to camera recorders, so in that way she was really helpful” [P8]. Several participants who had a fear of public speaking stated that they felt more comfortable rehearsing with the robot than with actual human audiences: “I prefer the robot, because with a [live] audience, you can see their expressions and that can be uncomfortable” [P11]. The presence of the robot as an attentive audience also helped the presenters “practice maintaining eye contact” [P12] and forced them to look away from their notes, further requiring them to internalize their talking points. The physical embodiment of the robot, and its ability to track presenters with its gaze, make this function particularly effective with a robot compared to other media. As a result, 11 of 12 participants said they would prefer rehearsing and receiving feedback from the robot over receiving the feedback in an audio-only format.
As recommendations for improvement, two participants suggested adding more human-like characteristics to the robot, both in terms of physical appearance and non-verbal behavior. Another participant suggested that the robot’s facial expressions be improved to create an impression of a “more friendly” coach [P3], which could help increase the presenter’s acceptance of negative feedback.

**Improving Presentation Quality Through Rich, Contextual Feedback:**

Most participants commended the usefulness of the coach’s “rich, customized and instant” [P7] feedback in helping them identify specific aspects of their presentations that they could act upon to improve their final delivery: “By practicing with the robot, I felt that I did much better because I would know what to work on between each take. While here I was confident in what I was doing, but I could have been totally wrong” [P4]. Participants reported varying opinions on which of the five feedback categories was the most helpful to them, but all agreed that they were important aspects of presentations. They also appreciated the high-level expert advice embedded in the introductory dialogues and the feedback: “She also gave really good tips that really helped me a lot, like, how you should give a presentation and how you should know your audience” [P1].

On the topic of feedback modalities, six participants stated that they would prefer the verbal feedback over graphical displays, due to its “informal conversation” [P2] style and its readily understandable nature. Other participants either expressed no preferences or suggested supplementing verbal feedback with more detailed visualizations for measures that might benefit from access to fine-grained data, such as pitch range.

To increase the applicability of the coach’s suggestions, several participants recommended including more “specific examples” [P12] in the feedback, or have the robot act as a role model to demonstrate good presentation techniques.

**Influencing Speaker Confidence:**
Participants reported mixed opinions regarding the effects of our coaching feedback on the speaker’s confidence. Three participants stated that the coach could help reduce public speaking anxiety because: “you are speaking to a robot and not people, so it would remove some stage fear. And it would correct you so you don’t make mistakes in public” [P10]. 7 of 12 participants highlighted the positive impact of our “feedback sandwich” strategy on boosting their motivation and confidence: “The fact that she was there to help me. It helped me a lot with each slide, boosting up my confidence after listening to all of the good points, and also listening to the feedback in case I have to improve” [P8]. Moreover, emphasizing positive performance trends also helped presenters feel more confident through a sense of improvement: “She was really good in that I made a mistake in the last time, I corrected it and then she would say ‘you really improved this from the last time’” [P1].

On the other hand, several participants reported increased anxiety due to constant reminders of needs for improvement from the robot. Some of them referred to this as a “good form of nervousness” [P5] as it encouraged them to “brush up a little more” [P5] on their performance. However, this could also have a detrimental effect on the confidence level for some participants, especially when they failed to make any noticeable improvement: “I thought I really worked on my pitch range, but she kept saying ‘you should still work on it’...It really killed my confidence” [P3].

Thus, further research is required to determine the appropriate frequency and timing of the coach’s suggestions. In addition, future systems should also incorporate mechanisms to dynamically set achievable goals based on the presenter’s characteristics and performance level.

7.5 Conclusion

In this chapter, I described the design and evaluation of RoboCOP, a fully-automated robotic presentation rehearsal coach and examined the effects of integrating presentation tracking in this system to provide feedback on content coverage. Compared to rehearsing alone and other non-
interactive forms of training, RoboCOP created a more engaging rehearsal environment that simulated a realistic presentation scenario with an attentive audience. While rehearsal for an audience is a recommended practice, not every speaker has easy access to a human coach or a knowledgeable listener who can give constructive feedback. Our robotic coach aimed to address this problem by providing detailed, actionable and empathetic feedback that resembled the behavior of human coaches.

Results of the feedback modality study showed that RoboCOP led to improvements in the rehearsal experience of presenters compared to graphical feedback and verbal feedback without the robot. Participants in our second evaluation study who rehearsed with RoboCOP reported very high levels of satisfaction with the system and desire to use it again for future rehearsals. Judges also rated RoboCOP-assisted presentations as significantly more engaging, novel, and exciting, and significantly better on overall presentation quality and presenter speech quality compared to non-assisted presentations. The content-based feedback generated from presentation tracking resulted in improvements in content coverage by the presenters.

The evaluation study has several limitations, beyond the small convenience sample of presenters and judges we used. Giving a final talk in front of an experimenter and video camera in a laboratory may be a poor proxy for real presentations; improvements made using RoboCOP may not actually carry over into real situations. The presentations we used were also very short and fully prepared, so they may not be representative of more typical talk preparation scenarios.
CHAPTER 8

Conclusion

In this dissertation, I developed a novel framework for real-time tracking of oral presentations and demonstrated its effectiveness in providing assistance for different presentation activities. I began by exploring the current practices and shortcomings in presentation assistance solutions and identified the lack of studies on providing assistance using presentation tracking. I designed a real-time oral presentation tracking framework using semantic matching of speech and presentation content and evaluated the effectiveness of this framework in different application scenarios. I focused on 3 different applications of presentation tracking: intelligent teleprompter systems, automatic slide recommendation to support dynamic presentations, and feedback on presentation content coverage during rehearsal.

8.1 Contributions

Particularly, this dissertation answered the following research questions:

1. How can speech recognition and semantic text retrieval techniques that have been successfully applied in similar applications be adopted to develop a real-time oral presentation tracking framework?

To answer this question, in Chapter 4, I described the design and evaluation of a presentation tracking framework. Based on my requirement analysis for a tracking framework, I first developed and evaluated a prototype system that tracks the presentation content coverage using ASR to spot manually extracted keywords in the presenter’s speech. The results of evaluation of this preliminary
implementation demonstrated the possibility and usefulness of developing a tracking framework. Based on findings from this formative study, I designed a real-time presentation tracking framework that automatically extracts and expands slide content keywords, assigns weights to them, and recognizes them in ASR output. I implemented this framework using techniques explored in my review of speech recognition and semantic text retrieval methods, and evaluated the accuracy of presentation tracking using different ASR outputs and keyword weighting methods. The results showed that ASR confusion network and keyword weighting using $tf.idf$ and semantic specificity resulted in the highest F-1 value of 76.9%. The technical implementation and evaluation of this framework is published in [3].

2. To what extent does an automatic teleprompter that utilizes presentation tracking improve the presenters’ experience?

Effective display of speaker notes can reduce the fear of forgetting the main speaking points and thus improve the presenter’s experience during delivery. To assess the effectiveness of presentation tracking in supporting intelligent note display, I developed and evaluated a speech-based, fully-automated teleprompter system called IntelliPrompter. As described in Chapter 5, this system uses presentation tracking to automatically identify the next speaking points and highlights them in slide notes. To explore different display modalities, I also examined displaying speaking notes on a head-mounted display. In a within-subjects user study with 30 participants comparing dynamic note display using different modalities with a traditional static note display system, participants expressed significantly higher satisfaction for the screen-based IntelliPrompter compared to the static note display, demonstrating the effectiveness of presentation tracking in improving presenters’ experiences. Presenters reported mixed feedback on head-mounted display modality, and judges rated the quality of presentation delivered using this modality significantly worse. On average, the system achieved 75.6% accuracy in tracking. Results of this study are published in [4].
3. When would presentation tracking be effective in supporting dynamic presentations?

In my exploratory study, QA sessions were identified as one of the most common situations in which dynamic presentation and non-linear navigation could be beneficial. In Chapter 6, I demonstrated the effectiveness of presentation tracking in providing support during QA sessions. I described the development and evaluation of Quester, a speech-based content recommendation system that utilizes the tracking framework to identify the presentation content related to presenter’s speech. During the presentation, Quester dynamically orders presentation slides based on their relatedness to speech and displays them at the bottom section of the presenter. It also highlights segments of slide notes that are related to speech. To improve the robustness of the system, I expanded the tracking framework with a speech content recency model. I evaluated this system in the setting of question-answering sessions at the end of presentations, which are a common scenario for dynamic presentations according to one of my exploratory studies. I conducted a within-subjects user study with 16 participants comparing Quester with a static slide navigation system. The participants expressed strong preferences for Quester, and performed significantly better during question-answering sessions. The average accuracy of the system in recommending correct slides was 84%. Results of this study are published in [5].

4. How effective is presentation tracking in providing content-based feedback during presentation rehearsal?

In Chapter 7, I examined the application of presentation tracking in presentation rehearsal. I integrated the tracking framework in RoboCOP, a robotic presentation rehearsal coach which simulates a training platform with attentive audience. Presentation tracking enables RoboCOP to provide feedback on presentation content coverage in addition to prosodic features of speech and eye contact with audience. Presentation rehearsal with RoboCOP resulted in significant increase in
quality of presentations and improved the content coverage by presenters compared to rehearsing without it. Results of this study are published in [110].

Based on the results of these studies, I conclude that utilizing presentation tracking in presentation assistance applications can improve the experience and the performance of presenters during rehearsal and delivery of presentations and QA sessions. Figure 31 provides a summary of the contributions of this dissertation and Table 11 summarizes the primary outcomes of the studies.

Figure 31. Summary of the contributions of this dissertation
### Table 11. Primary outcomes of the studies

<table>
<thead>
<tr>
<th>Study Name</th>
<th>Primary Outcomes</th>
</tr>
</thead>
</table>
| IntelliPrompter | Average tracking accuracy of 75.65%  
Significant increase in usability ratings (17.67% increase), compared to a static note display system |
| Quester | Average slide recommendation accuracy of 84.37%  
Significant increase in usability ratings (34.95% increase), compared to a system that only provides linear slide navigation  
Significant decrease in delay for QA (48.65% decrease) |
| RoboCOP | Average tracking accuracy of 82.62%  
Significant increase in audience satisfaction with presenters (12.10% increase), compared to rehearsing without RoboCOP |

### 8.2 Limitations

#### 8.2.1 The Note Authoring Process

The availability and form of speaking notes is a key aspect in presentation tracking. Future studies could benefit from a more in-depth examination of the note authoring process:

*Availability of Speaking Notes*

The tracking framework uses slide notes as the main source of content for tracking. However, my analysis of a presentation notes corpus, reported in section 5.1, showed that only 21% of presentations contained notes. One suggestion for eliminating the need for authoring notes is to use the presenter’s speech transcript during presentation rehearsal. This can also potentially encourage spoken presentation rehearsal, which is a recommended practice [70].

Furthermore, additional sources of information such as web and local documents can be utilized to support dynamic presentations. The framework can suggest supporting material related to
presenter’s speech in real-time. Some examples of extra content include: related publications, multimedia sources, and other presentations. Suggestion of content during presentation authoring have been examined in [61].

The Form of Speaking Notes

Although the form of speaking notes (e.g., full sentences, keywords, bullet points) does not affect the performance of the tracking system (since the system automatically extracts and uses the keywords), it might influence the user’s experience and the overall effectiveness of the system. In conducted studies, the slides notes only contained prose text, but other common forms of notes such as key words or bullet points should also be examined.

In addition, as mentioned in section 4.3.1, presentation content segmentation is an important factor in the performance of the tracking system. In the current framework, the slide content was manually segmented into topics, considering the length, subject and number of keywords in each topic. Automating the segmentation process can simplify the process of authoring of slide notes for tracking. Automatic topic segmentation has been the focus of several studies in natural language processing [27, 30, 48].

Dependency between Slides and Slide Notes

As an alternative approach to topic-based content segmentation, the presentation tracking framework can take advantage of the dependency between slide visuals and note sections to automatically structure speaking notes into individual segments corresponding to specific visual elements on the slide, similar to the element note authoring supported in the PitchPerfect system [112].

This dependency can also be used to automatically modify the appearance of slides for the audience based on the presenter’s speech. For example, it can highlight visual and textual elements of slides
that are related to covered segments of presentation. This provides an automatic visual indicator of the presentation progress to the audience without reliance on specific order of presentation.

8.2.2 Need for More Comprehensive Evaluations

The evaluation studies conducted in this dissertation used small convenience samples of speakers and thus are likely not representative of all speakers. Thus, more comprehensive studies with a diversity of participants and presentation subjects are required to further evaluate the effectiveness of the tracking framework and its applications.

Diversity of Speakers

The performance and reliability of the ASR system and consequently the tracking system is highly dependent on the speaker, therefore evaluating the system with a wider range of participants with different accents, speech abilities (including those with speech impairments) and language proficiencies can provide a more generalized assessment [10]. In addition, the perceived helpfulness and ease of use of the systems could also be affected by the professional and educational background and the level of computer literacy of the user.

Presentation Subject

My evaluation studies included presentations with general subjects that were relatively interesting and did not require any specialized knowledge from the participants. Although the tracking framework is designed to be independent of the subject, it could potentially have a lower accuracy for presentations with less frequent words, since the ASR might have difficulties detecting those words [9]. The performance and effectiveness of the framework should be evaluated with more technical and scientific presentations and more diverse subjects.
Realistic Settings

The studies were conducted in controlled environments and conditions. Factors such as quality of audio recording and noise can affect the performance of the system [57]. Furthermore, users may have different opinions regarding the usefulness of the system in more stressful and critical situations. Evaluating the system in realistic settings such as classrooms lectures and conference presentations can result in deeper insights into the practicality of presentation tracking.

8.3 Future Work

8.3.1 Improving the Performance of the Tracking Framework

In the evaluation studies presented in this dissertation, the tracking framework achieved an average accuracy of 76.9%. Although this level of performance was shown to be adequate for improving the experience of presenters and in some cases their performance, further enhancement can potentially result in even better results. Some suggestions for improving the performance of the system include:

Structure of Speech

Information about the structure of speech can be exploited to improve the accuracy of content tracking, as shown previously in off-line speech-to-text alignment applications [63]. Previous studies have explored automatic topic segmentation in speech using prosodic information [42, 100].

Composition of Presentation

Inspired by previous studies [32, 60], the interaction between the tracking framework and the presenter can be described as a collaborative process, with the user dynamically initiating presentation about different topics and the framework displaying the proper slides. Plan recognition models [20] can be used for human-agent collaboration by exploiting hierarchical task plans [56].
In a similar approach, the information about the composition of the presentation and the precedence relationship of slides can be utilized to determine the set of slides with the highest probability of being presented at each moment. This hierarchical structural information can be provided by the user in addition to slide notes and may include the sequence of slides in each section and subsection of the presentation.

8.3.2 Alternative Applications

In this dissertation, the tracking framework was evaluated in 3 different applications related to presentation rehearsal and delivery, but it can be utilized and assessed in many other situations. Some suggested alternative applications and settings include:

*Extemporaneous Talks*

In Chapter 6, I evaluated the effectiveness of tracking in assisting presenters during QA sessions. As suggested by some participants in our exploratory study, review lectures and extemporaneous talks are another setting for dynamic presentations. Using the tracking framework, the system could recommend related slides based on the speech content, and the presenters could use the system to dynamically tailor their delivery based on their real-time evaluation of the audience’s level of understanding, as well as direct input or questions from the audience.

*Other Forms of Visual Support*

The applications of presentation tracking can be extended to other forms of visual support such as canvas presentations [58], video voice-overs, and live demos. For instance, presentation tracking can be used for automatic navigation through zoomable user interfaces [40] based on the presenter’s speech during presentation delivery. Alternatively, it can encourage interactivity in traditionally linear types of presentations such as video voice-overs by playing segments of a video that are related to audience feedback.
**Meetings**

Beyond the context of oral presentations, tracking can also be used to facilitate meetings. Using the tracking framework, a meeting support system can automatically suggest related supporting material based on the topic currently being discussed. It can also track the meeting progress against a pre-defined meeting agenda to highlight which items have been covered and which are remaining. Compared to presentations, meetings have a less strict structure and more dynamic flow. In addition, meetings include multiple speakers which can make tracking more challenging.
Appendix A

General Questionnaires Used in All User Studies

A.1 Sociodemographics Questionnaire

Please take a moment and answer a few questions about yourself:

Date of Birth: _________

Sex:  M / F

Ethnic Background (check one):

- American Indian or Alaskan Native     ___
- Asian or Pacific Islander             ___
- Black, Not of Hispanic Origin         ___
- White, Not of Hispanic Origin         ___
- Hispanic                               ___

Last grade of school completed (check one):

- Less than high school (0-8)           ___
- Some high school                      ___
- High school graduate or GED           ___
- Technical school education            ___
- Some college                          ___
- College graduate                      ___
- Advanced degree                       ___
Occupation: __________________________________________

How often do you read books (check one):

Never ______
Less than once a week ______
Once a week ______
A few times a week ______
Every day ______

How much experience do you have with computers (check one):

I’ve never used one. ______
I’ve tried one a few times. ______
I use one regularly. ______
I’m an expert. ______

How much experience do you have giving oral presentations (check one):

I’ve never given one. ______
I’ve given a few. ______
I’ve given many. ______
I am an expert presenter. ______

How much experience do you have using PowerPoint (or similar) (check one):

I’ve never used one. ______
I’ve tried one a few times. ______
I use one regularly. ______
I’m an expert. ______
A.2 Self-Perceived Communication Competence Scale (SPCC)

Reference: [69]

Directions: Below are twelve situations in which you might need to communicate. People's abilities to communicate effectively vary a lot, and sometimes the same person is more competent to communicate in one situation than in another. Please indicate how competent you are to communicate in each of the situations described below. Indicate in the space provided at the left of each item your estimate of your competence.

Presume 0 = completely incompetent and 100 = competent.

_____1. Present a talk to a group of strangers.
_____2. Talk with an acquaintance.
_____3. Talk in a large meeting of friends.
_____4. Talk in a small group of strangers.
_____5. Talk with a friend.
_____6. Talk in a large meeting of acquaintances.
_____7. Talk with a stranger.
_____8. Present a talk to a group of friends.
_____9. Talk in a small group of acquaintances.
_____10. Talk in a large meeting of strangers.
_____11. Talk in a small group of friends.
_____12. Present a talk to a group of acquaintances.
A.3 Personal Report of Confidence as a Speaker (PRCS)

Reference: [83]

Directions: This instrument is composed of 30 statements regarding your feelings of confidence as a speaker. For each statement, there is a “true” and a “false” option. Please try to decide whether “true” or “false” most represents your feelings associated with your most recent speech, then write down either “true” (T) or “false” (F) next to each statement. Please work quickly and don’t spend much time on any question. We want your first impression on this questionnaire. Now please go ahead, work quickly, and remember to answer every question.

1. _____ I look forward to an opportunity to speak in public.
2. _____ My hands tremble when I try to handle objects on the platform.
3. _____ I am in constant fear of forgetting my speech.
4. _____ Audiences seem friendly when I address them.
5. _____ While preparing a speech I am in a constant state of anxiety.
6. _____ At the conclusion of a speech I feel that I have had a pleasant experience.
7. _____ I dislike to use my body and voice expressively.
8. _____ My thoughts become confused and jumbled when I speak before an audience.
9. _____ I have no fear of facing an audience.
10. _____ Although I am nervous just before getting up, I soon forget my fears and enjoy the experience.
11. _____ I face the prospect of making a speech with complete confidence.
12. _____ I feel that I am in complete possession of myself while speaking.
13. _____ I prefer to have notes on the platform in case I forget my speech.
14. _____ I like to observe the reactions of my audience to my speech.
15. _____ Although I talk fluently with friends I am at a loss for words on the platform.
16. _____ I feel relaxed and comfortable while speaking.
17. _____ Although I do not enjoy speaking in public, I do not particularly dread it.
18. _____ I always avoid speaking in public if possible.
19. _____ The faces of my audience are blurred when I look at them.
20. _____ I feel disgusted with myself after trying to address a group of people.
21. _____ I enjoy preparing a talk.
22. _____ My mind is clear when I face an audience.
23. _____ I am fairly fluent.
24. _____ I perspire and tremble just before getting up to speak.
25. _____ My posture feels strained and unnatural.
26. _____ I am fearful and tense all the while I am speaking before a group of people.
27. _____ I find the prospect of speaking mildly pleasant.
28. _____ It is difficult for me to search my mind calmly for the right words to express my thoughts.
29. _____ I am terrified at the thought of speaking before a group of people.
30. _____ I have a feeling of alertness in facing an audience.
A.4 State Anxiety Questionnaire

Reference: [105]

A number of statements which people have used to describe themselves are given below. Reach each statement and then circle the appropriate number to the below the statement to indicate how you feel right now, that is, at this moment. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

I feel calm
- Not at all……Somewhat ......Moderately So......Very much So

I feel secure
- Not at all……Somewhat ......Moderately So......Very much So

I feel tense
- Not at all……Somewhat ......Moderately So......Very much So

I feel strained
- Not at all……Somewhat ......Moderately So......Very much So

I feel at ease
- Not at all……Somewhat ......Moderately So......Very much So

I feel upset
- Not at all……Somewhat ......Moderately So......Very much So

I am presently worrying over possible misfortunes
- Not at all……Somewhat ......Moderately So......Very much So

I feel satisfied
- Not at all……Somewhat ......Moderately So......Very much So

I feel frightened
- Not at all……Somewhat ......Moderately So......Very much So

I feel comfortable
<table>
<thead>
<tr>
<th>Feeling</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel self-confident</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel nervous</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel jittery</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel indecisive</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel relaxed</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel content</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel worried</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel confused</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel steady</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
<tr>
<td>I feel pleasant</td>
<td>Not at all……Somewhat</td>
</tr>
<tr>
<td></td>
<td>………Moderately So……Very much So</td>
</tr>
</tbody>
</table>
Appendix B

Study Protocols for IntelliPrompter

Study

B.1 Study Protocol for Presentation Sessions

[Before participant comes]

a. Change Participant ID and First Name in the participant.xml file

b. Make sure the Notes file has been generated beforehand

c. Make sure microphone and camera are charged

1. Thank you for coming in today to help us out with our research study. Let me first just double check that you are eligible. Are you at least 18 years old? And able to speak and read English? Do you have college education? Do you have experience giving presentations? [If no, thank and dismiss, if yes continue.]

2. Let me tell you a little bit about what we will be doing in this study. We’re exploring how technologies can be used to display speaking notes while people deliver oral presentations. In this session, we will give you a pre-made PowerPoint slide deck and notes, and give you 30 minutes to prepare. You will then deliver your presentation with the help from different teleprompter systems. Your presentation will be videotaped. The session should take about 90 minutes, and you will be compensated $25 at the end of the session.

3. Does this sound like something you are still interested in? [If yes, continue, if no, dismiss]
4. Ok. Great. Now the first thing we will do is going over a consent form. I will explain the form to you and you can take as much time as you want to read it and ask any questions you may have. [Administer informed consent]

5. Now, I would like you to fill out a short questionnaire providing some information about yourself. Let me know when you are finished. [Hand participant socio-demographic]

6. The next questionnaire is going to ask you some questions about how you perceive your communication competence. [Self-perceived Communication Competence Scale]

7. Great, thank you. The next question is going to ask you some questions about your feelings of confidence as a speaker. These questions ask about your feelings associated with your most recent presentation [Personal Report of Confidence as a Speaker]

8. Ok, now we can get started. I would like you to give a 9-minute presentation on Italy. Your audience will be a group of tourists are who considering to take a trip to Italy. Your goal is to introduce some of the most interesting information about Italy, and to convince them to take a trip to Italy.

We will break your presentation into 3 parts. Each part should be 3-minute long. In each part, you will talk about one aspect of Italy that may attract tourists. You will rehearse, and then deliver each part separately, with the help of different teleprompter systems that display your speaking notes while you are speaking.

We have already prepared the PowerPoint slides and speaking notes for each of the three parts. You should try to cover all the main points in the notes, but not necessarily word-for-word.

9. [Open First Part slides] Now let’s take a look at the slides for the first part of your talk, about [Tourist Destinations of Italy/ Italian Art/Italian Cuisine]. There are 3 slides for the first part. You will receive a different set of slides when we go to the second part. For now, let's just focus on the first part. Please do not change the slides.
10. [Show participants the notes] You can see the notes for each slide in this panel [show Notes pane for Slide 1]. The notes are already divided into main points. If you click on the title of each point, you will see more details about it. Please do not change the notes [Prompt participant to try clicking on a point]. You will be able to see these notes when you practice and deliver your talk.

11. Great. Now, we will give you 5 minutes to review the slides and notes.
    a. After 5 minutes, we will ask you to practice this part. We won't be in the room while you rehearse.
    b. After your rehearsal, you will present the first part of your talk. We will be in the room when you present, and your talk will also be videotaped.
    c. After this study, your final presentation will be shown to a panel of judges who will evaluate the presentation quality. There are two main criteria that the judges will use to evaluate your presentation. First is Content Coverage (i.e. how many topics in the notes that you manage to cover in your presentation) Second is Note Reliance (i.e. how often do you have to look at the notes while presenting). So please try not to just read off the notes.
    d. OK, now please take your time to review your slides and notes. I will come in after about 5 minutes. If you are ready before that, just let me know – I will just be waiting outside.
    e. [If participant asks for papers to take notes] Part of our study is to see how technology can support people who don't take notes on paper, so we would prefer you not using paper notes.
    f. [If participant asks for more time] Unfortunately, as part of the study we need to stay on schedule, so we're afraid we won't be able to give you more time. But you still have time now to rehearse.

12. [After 5 minutes, start Presentation] OK, it's time to start your rehearsal.
    a. In this rehearsal, we would like you to go through and practice presenting each slide. While you practice, we would like you to stand up and speak aloud, as if you are presenting in front of an audience.
    b. Here is the remote control for you to advance the slides while you present. You can use the Right Arrow button to go to next slide, and Left Arrow button to go to previous slide. [Ask them to go to the first slide]
c. Also, as part of the study, we ask you to wear this Google glass device while you are speaking. This is a head-mounted display that has a small screen to show information.

[Control/Screen] For this part, it won't display any info. But later on, in another part of the presentation, you will see your notes displayed on that screen.

d. You will see all the notes displayed on the notes pane here [point to the notes]. The first topic that you are supposed to present is highlighted and displayed at the top of the notes.

e. [Control Condition]: While you are speaking, if you want, you can use the remote control to manually navigate to the topic that you want to say next. That topic will be highlighted and moved to the top of the notes. You can use the Up Arrow button to move up the topic list, and the Down Arrow button to move down the topic list.

f. [Screen Condition]: While you are presenting, the system is going to listen to what you are saying. Based on your speech, it will automatically detect what topic that you have covered and what topic you are likely to say next. The topics that you have covered will be marked with a green tick. The next topic will be highlighted and moved to the top of the notes.

If you want, you can also use the remote control to navigate to the topic that you want to say next. You can use the Up Arrow button to move up the topic list, and the Down Arrow button to move down the topic list.

g. [Google Glass condition – Start the presentation mode on Glass, make sure the participant sees the Welcome message]: Please adjust the screen so that you can see the whole text clearly.

While you are presenting, the system is going to listen to what you are saying. Based on your speech, it will automatically detect what topic that you have covered and what topic you are likely to say next. The topics that you have covered will be marked with a green tick, and the teleprompter will transition to the next topic.
If you want, you can also use the remote control to manually navigate to the topic that you want to say next. You can use the Up Arrow button to move up the topic list, and the Down Arrow button to move down the topic list.

h. Whenever you are ready, just press the "Presentation" button [Point to the button]. When you are finished, please come out and let me know. I will just be waiting outside.

13. [End of rehearsal, before delivery] Now it is time for you to present the first part. But before you start your presentation, I just need you to fill in a questionnaire about how you feel right now [State Anxiety]

14. Great. As I mentioned before, we will videotape your final presentation. This presentation will be shown to a panel of judges after the study. The judges will evaluate your presentation on Content Coverage (i.e. how many topics in the notes that you manage to cover in your presentation), and Note Reliance (i.e. how often do you need to look at the notes instead of looking at the audience). If you want to introduce yourself in the presentation, please introduce yourself by your first name only. For confidentiality, we don’t want your full name to appear in the presentation video.

15. Whenever you are ready, please press the "Present" button. [Press Record button when ready]

16. [After presentation] Great. Thank you for giving the presentation. Now I would like you to fill out a questionnaire. You have filled in this questionnaire once at the beginning of the session. This questionnaire asks about your feelings associated with your most recent presentation. In this case, it refers to the presentation you have just given [Personal Report of Confidence as Speaker]

17. Great. Now I would like you to fill in a questionnaire about your experience with the teleprompter system that you used during your presentation [Teleprompter System Rating]
18. Great. Now I just need to you fill in a questionnaire about your evaluation of the talk that you just gave [Presentation Quality Rating]

19. [Second condition: Open the Second Part slides] Now let’s move on to the second part of your presentation. In this part you will be talking about [Tourist Destinations of France/ French Art/French Cuisine] [Repeat steps 11-18]

20. [Third condition: Open the Third Part slides] Now let’s move on to the final part of your presentation. [Tourist Destinations of Italy/ Italian Art/ Italian Cuisine] [Repeat steps 11-18]

21. [After all conditions] Now I would like to ask you to rank the three teleprompter systems that you have used. In this questionnaire, please rank the teleprompter systems based on each of these criteria, from 1st (Best) to 3rd (Worst). [Ranking questionnaire]

22. That concludes the computer portion of the study. Lastly, I would just like to ask you a few questions about your experience with the tasks. [Conduct semi-structured interview, use phone to record, refer to interview sheet]
B.2 Interview Guide

1. How often do you give presentations? How often do you write speaking notes for your presentations?
   a. Have you encountered any difficulties using notes while presenting in the past?

2. How do you feel about the three parts of the talk that you gave today?
   a. Which part do you think you did the best, and why? Which part do you think needs the most improvement, and why?
   b. How well do you think you did compared to your previous/similar presentations?

3. Of the three teleprompter systems that you used today, which one do you like the most? Why?
   a. How do you think the system helped you? (in terms of knowing what to say next, note reliance, eye contact, content coverage, confidence boost, presentation quality)
   b. What did you like least about the system?
   c. What did you like most about the system?

4. Of the three teleprompter systems that you used today, which one do you like the least? Why?
   a. How do you think the system helped you, in any ways?
   b. What did you like least about the system?
   c. What did you like most about the system?

5. How would you compare the [2nd ranked] system with the [1st ranked] system? With the [3rd ranked] system?

6. [Screen/Glass conditions] How accurate was the system in terms of detecting what topics have been covered and what topic to say next?

7. How would you compare between presenting with the Google Glass vs. without the Glass (e.g., comfort, distraction)? Which one would you prefer?

8. Would you use any of these teleprompter systems in the future? Why/why not?
   a. Are there other features that you would like to add to these systems?
B.3 Protocol for Judge Study

1. Thank you for coming in today to help us out with our research study. Let me first just double check that you are eligible. Are you at least 18 years old? Able to speak and read English? Have some college education? Have experience giving presentations? [If no, thank and dismiss, if yes continue.]

2. Let me tell you a little bit about what we will be doing in this study. We are developing a new technology that helps people deliver oral presentations. In this study, we want to evaluate the quality of presentations given using our technology. We will ask you to view a series of short videotapes of oral presentations. After each viewing, we will ask you to fill out questionnaires to measure your assessment of the quality of the presentation. The study takes approximately 60 minutes, and you will be compensated $15 at the end of the session.

3. Does this sound like something you are still interested in? [If yes, continue, if no, dismiss]

4. Ok. Great. Now the first thing we will do is going over a consent form. I will explain it to you and you can take as much time as you want to read it and ask any questions you may have.

5. Now, I would like you to fill out a short questionnaire providing some information about yourself. Let me know when you are finished. [Hand participant socio-demographic]

6. The next questionnaire is going to ask you some questions about how you perceive your communication competence. [Self-perceived Communication Competence Scale]

7. Now we can get started. I would like you to watch a video of an oral presentation. In this presentation, the presenter was asked to give a 3-minute presentation on [Tourist Attractions of Italy/Italian Art/Italian Cuisine]. Their target audience is a group of tourists who are considering to take a trip to Italy. Their goal is to introduce some interesting facts about [Attractions/Art/Cuisine], and to convince their audience to take a trip to Italy. After you
viewed the presentation, we will ask you some questions about your assessment of the quality of the presentation.

8. **[After viewing 1st presentation]** Now, I would like you to fill out a questionnaire about the presentation you just viewed. This questionnaire has two parts. The first part asks about your evaluation of the **presentation quality**. The second part of the questionnaire asks about your evaluation of the **presenter** in this presentation **[Presentation Rating]**.

9. **[Second presentation]** I would like you to view a second presentation. In this presentation, the presenter was also asked to give a **3-minute** presentation, but on the topic of **[Attractions/Art/Cuisine]**. Similar to the first presentation, their target audience is a group of tourists who are considering to take a trip to Italy. Their goal is to introduce some interesting facts about **[Attractions/Art/Cuisine]**, and to convince their audience to take a trip to Italy. After you viewed the presentation, we will again ask you some questions about your assessment of the quality of the presentation.

10. **[After viewing 2nd presentation]** Now I would like you to fill in this questionnaire about the quality of the presentation you just viewed, and the presenter in this presentation **[Presentation Rating]**.

11. **[Compare 1st and 2nd presentations]** OK. So, you have watched two presentations. The first presentation is about **[Tourist Attractions/Art/Cuisine]**, the second presentation is about **[Tourist Attractions/Art/Cuisine]**. Now we would like you to fill in a questionnaire comparing the quality of the two presentations you just viewed.

   a. We would like you to compare the two presentations on five criteria: organization, content coverage, note reliance, speech, and overall quality.

   b. Now let me briefly go through the criteria with you:
i. Organization: whether information was presented in a logical sequence, whether the presenters make smooth transitions from point to point and slide to slide.

ii. Content coverage: Whether presenters covered the main points in sufficient details.

iii. Note reliance: How often the presenters needed to refer to the notes? Did they maintain good eye contact with the audience?

iv. Speech: How often did they use distracting pauses and fillers?

v. Pacing & Timing: Did they speak too fast or too slowly? Were the presentations too long, too short, or at the right length? (remember that the presenters were asked to give a 3-minute presentation)

c. For each criterion, please indicate if there is any difference between the two presentations, and the level of differences (no difference, slight difference, moderate difference, substantial difference). Then indicate which presentation is better (first or second).

12. [Third presentation] Now I would like you to view a third presentation from the same presenter. In this presentation, the presenter was also asked to give a 3-minute presentation, but on the topic of [Attractions/Art/Cuisine]. After you viewed the presentation, we will again ask you some questions about your assessment of the quality of the presentation.

13. [After viewing 3rd presentation] Now I would like you to fill in this questionnaire about the quality of the presentation you just viewed, and the presenter in this presentation [Presentation Rating].
14. [Compare 2nd and 3rd presentations] Now we would like you to compare the quality of the last two presentations you just viewed. The second presentation you viewed is about [Tourist Attractions/Art/Cuisine], and the third one that you viewed just now is about [Tourist Attractions/Art/Cuisine].

   a. We would like you to compare the two presentations on five criteria, as before: organization, content coverage, note reliance, speech, and overall quality.

   b. For each criterion, please indicate if there is any difference between the two presentations, and the level of differences (no difference, slight difference, moderate difference, substantial difference). Then indicate which presentation is better (second or third).

15. [Review 1st presentation] Now we would like you to compare the quality of the third presentation and the very first presentation you viewed. To help you have an accurate comparison, we would like you to re-watch the first presentation.

16. [Compare 1st and 3rd presentations] Now we would like you to compare the quality of first presentation that you just reviewed and the third presentation.

   a. We would like you to compare the two presentations on five criteria, as before: organization, content coverage, note reliance, speech, and overall quality.

   b. For each criterion, please indicate if there is any difference between the two presentations, and the level of differences (no difference, slight difference, moderate difference, substantial difference). Then indicate which presentation is better (first or third).

17. [Second set of presentations] Now we would like to repeat the process we just did, but with another set of presentations from another presenter [repeat steps 6-11]
Appendix C

Questionnaires for IntelliPrompter Study

C.1 Teleprompter Rating Questionnaire

Please answer the following questions about the teleprompter system:

Write an ‘X’ on each line (on one of the dots):

How easy was it to use the teleprompter system?

Very difficult • • • • • • • • Very easy

How reliable was the teleprompter system?

Not at all • • • • • • • • Very much

How much do you feel the teleprompter system helped you?

Not at all • • • • • • • • Very much

How distracting was the teleprompter system?

Very distracting • • • • • • • • Not at all

How satisfied are you with the teleprompter system?

Not at all • • • • • • • • Very satisfied

How much would you like to give future presentations with the system?

Not at all • • • • • • • • Very much

With the teleprompter, it was always obvious what to say next.

Disagree • • • • • • • • Agree

I would feel comfortable wearing the display in front of an audience.

Disagree • • • • • • • • Agree
C.2 Presentation Self-Rating Questionnaire

Please answer the following questions about the presentation you just delivered:

*Write an ‘X’ on each line (on one of the dots):*

**How engaging** was your presentation?

| Not at all | • | • | • | • | • | • | • | Very much |

**How understandable** was your presentation?

| Not at all | • | • | • | • | • | • | • | Very well |

**How nervous** were you during your presentation?

| Not at all | • | • | • | • | • | • | • | Very much |

**How exciting** was your presentation?

| Very boring | • | • | • | • | • | • | • | Very exciting |

**How entertaining** was your presentation?

| Not at all | • | • | • | • | • | • | • | Very much |

**How competent** were you during your presentation?

| Not at all | • | • | • | • | • | • | • | Very much |

How would you rate the **overall quality** of your presentation?

| Very poor | • | • | • | • | • | • | • | Very good |
### C.3 Ranking of the Teleprompter Systems

*Please rank the 3 different teleprompter systems based on each criterion from 1st (best) to 3rd (Worst):*

<table>
<thead>
<tr>
<th>System A:</th>
<th>System B:</th>
<th>System C:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Criteria</th>
<th>1st (Best)</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of use of the system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability of the system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helpfulness of the system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The level of distraction caused by the system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The system that made it obvious what to say next</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with the system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desire to continue using the system</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C.4 Judge Presentation Rating Questionnaire

Please answer the following questions about the presentation you just viewed:

Write an ‘X’ on each line (on one of the dots):

**Evaluation of the Overall Presentation:**

<table>
<thead>
<tr>
<th></th>
<th>Very poor</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very good</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you rate the overall quality of the presentation?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very well</th>
</tr>
</thead>
<tbody>
<tr>
<td>How well could you understand the presentation?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much were you engaged by the presentation?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Very routine</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>How novel was the presentation?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Very boring</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very exciting</th>
</tr>
</thead>
<tbody>
<tr>
<td>How exciting was the presentation?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>How entertaining was the presentation?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much would you like to see another presentation like this?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Evaluation of the Presenter:**

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>How satisfied are you with the presenter?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How competent was the presenter?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How nervous was the presenter?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How well could you understand the presenter?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How engaging was the presenter?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How exciting was the presenter?</td>
<td>Very boring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How entertaining was the presenter?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### C.5 Judge Presentation Comparison Questionnaire

<table>
<thead>
<tr>
<th>Criteria</th>
<th>No Difference</th>
<th>Slight Difference</th>
<th>Moderate Difference</th>
<th>Substantial Difference</th>
<th>Which presentation is better, if any?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ORGANIZATION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logical sequence of information, transitions from point to point and slide to slide</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td><strong>CONTENT COVERAGE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amount of points covered in sufficient depth</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td><strong>NOTE RELIANCE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of reference to notes during delivery</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td><strong>SPEECH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Articulation, frequency of distracting pauses and fillers (e.g. ‘uh’, ‘um’), vocal variety, speaking rate</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td><strong>TIMING &amp; PACING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacing per slide and overall length of the presentation</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td><strong>OVERALL QUALITY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Slide 1:
Italian art has influenced several major movements and has produced some of the greatest artists of all time.

Italy has more than 400 art galleries, museums and exhibitions with around 40 million visitors every year.

The Vatican Museums have around 20000 artworks on display from the collection gathered by the Popes throughout the centuries.

The Uffizi Gallery in Florence is one of the largest museums in the world with priceless works from the Renaissance period.
Slide 2:

Leonardo Da Vinci, an Italian artist, scientist, and engineer, is renowned primarily as the greatest painter of all time.
His painting The Last Supper is the most reproduced religious painting of all time.
The Vitruvian Man is a drawing which is regarded as a cultural icon, and has been reproduced on the euro coin.
Michelangelo, from the 16th century, sculpted his best known work, David, before the age of thirty.
The Creation of Adam is a painting by him on the Sistine Chapel's ceiling. It illustrates the Biblical creation narrative.
Slide 3:

Italy plays a significant role in the history of European music. Many musical instruments were invented in Italy.

National Academy of Saint Cecilia, is one of the oldest musical institutions in the world which was founded in 16th century.

Birthplace of the opera, has some of the most famous opera houses in the world, including La Scala in Milan.

A museum adjacent to this opera house displays historic costumes and set designs.
D.2 Italy Tourist Attractions

Slide 1:

Italy is the fifth most visited country in the world with Rome, Milan, and Venice as the top destinations.

The Alpes mountains in the north and the Mediterranean Sea in the south attract tourists.

Italy has the greatest number of UNESCO World Heritage Sites (50), and has broad cultural influence.

Very diverse architectural style, with about 100000 monuments of all varieties. Has some of the finest works in Western architecture.
Rome is the capital of Italy and has more than 2500 years of history. It is the most popular tourist attraction in Italy.

Colosseum is the largest amphitheater in the world. It was built in the first century with the capacity of 70000 spectators.

Vatican City is Rome is home to religious sites such as Saint Peter's Basilica and the Sistine Chapel. The Trevi Fountain, from the 18th century, is one of the most famous fountains in the world.
Slide 3:

Venice is made up of 118 small islands that are separated by canals connected by bridges. It has been ranked the most beautiful city in the world. It was a major financial power during the Middle Ages.

Milan Cathedral is the largest cathedral in Italy and the fourth largest cathedral in the world. It is famous for its Gothic architectural style and took 6 centuries to build.

The freestanding bell tower of the cathedral of Pisa is known for its unintended tilt to one side.
D.3 Italian Cuisine

Slide 1:

One of the most popular in the world, with influences abroad. Its roots go back to 4th century BC. It has extreme simplicity; many dishes only have 4 to 8 ingredients. Tomatoes, garlic, olive oil, green vegetables and whole grains are the most common ingredients. Italian cooks rely on the quality of the ingredients rather than complex preparation. It has great regional diversity; each region has its own local ingredients and traditional recipes.
The modern pizza was invented in Naples, Italy. Pizza Margherita is the original variant of Italian pizza. Authentic versions must be made with a special type of tomatoes and buffalo mozzarella cheese. Famous for the diverse variety of pasta. There are hundreds of different shapes of pasta such as spaghetti and rigatoni.

Tiramisu (meaning "cheer me up") is a popular coffee flavoured Italian custard dessert. Gelato is the Italian style of ice cream which has less air and more flavoring which gives it a unique richness.
Slide 3:

Italy is the world's largest producer, exporter and consumer of wine. Italian wines have broad variety. Grapes for winemaking are grown in almost every region of the country and there are more than one million vineyards.

Italian style coffee, also known as espresso, is generally thicker than other coffees. It is base for other drinks.

Cappuccino is prepared with double espresso, hot milk, and steamed milk foam which can be decorated.
Appendix E

Study Protocols for Quester Study

E.1 Protocol for Formative Study

1. Thank you for helping us out with our research. Let me tell you a little bit about our project. We are developing new technologies that help people prepare and deliver their oral presentations. More specifically, we want to explore how technology can support a more dynamic form of presentations, where you can effectively adapt your delivery based on changes in audience responses and presentation contexts. To help design our system, we would like to learn about how expert presenters prepare and deliver oral presentations to audiences.

2. In this study, we will first show you a video clip of an oral presentation. We will then interview you about the scenario presented in the video clip, as well as your presentation experiences in general. At the end of the study, we will show you a demo of a presentation support prototype that we’ve made, and ask for your feedback on our prototype. The study should take around 45 minutes, and you will be compensated $15 at the end of the session.

3. Now I would like you to watch a very short vignette of a presentation. In this presentation, the presenter is talking about Italy in front of an audience. After you watch the video clip, I will ask you a few questions related to the scenario presented in the video. [Show video]

4. Now, I would like you to ask you a few questions about the video you just watched, and your experiences of preparing and delivering oral presentations. Would you mind if I record our interview? [Use phone to record, refer to interview guide]
E.2 Interview Guide for Formative Study

Dynamic Presentation Scenario

1. In this video clip, the presenter received a question from the audience in the middle of her talk. How well do you think she handled this situation?
   a. Would you do the same if you were in her situation? If not, how would you handle it?

2. Have you ever been in a similar situation when you need to dynamically tailor your delivery? Tell me about it.
   a. Tell me any other situations that you might need to dynamically tailor your delivery.

3. What are the main challenges you face when dealing with those situations? How do you overcome them?

General Info

1. Now let’s talk about your general presentation experiences. How often do you present?

2. What different types of presentations do you give?

3. How long do you usually talk for?

4. How large are your audiences?

Presentation Preparation

1. Now let’s talk about how you prepare for your presentations. How do you usually prepare for your presentations?

2. What are the main challenges you face when preparing?

3. How big is your slide deck usually?

4. When preparing a presentation, to what extent do you reuse past presentation material, in part or whole?

5. How do you prepare for those situations when you might need to dynamically tailor your delivery?
   a. Do you prepare backup materials, such as extra slides? If yes, how do you organize those materials? How useful are they to you?
b. If using extra slides, how big is your extra slide set compared to your main slide set? [Maybe ask them if they can give us some sample slide decks that have extra slides]

6. Do you write notes for your presentations?
   a. What types of notes do you prepare (e.g. full script, bullet points, keywords)?

Presentation Delivery & QA

1. Now let’s talk about your presentation delivery. What are the main challenge you face when delivering your presentations to your audiences?
   a. How do you overcome them?
2. How do you adapt your delivery to audience feedback and presentation contexts?
3. When presenting, to what extent are you prepared to present additional information to back up your points? How important is this for you?
4. If you do prepare backup materials, how do you use these materials during delivery?
   a. What problems do you have when using these materials during delivery (e.g. accessing and retrieving relevant information, navigating between backup and main materials?)
5. How do you prepare for QA sessions?
   a. What are the main challenges you face during QA sessions?
   b. How often do you use slides to help you answer audience’s questions? If yes, do you have any problems accessing and navigating to relevant slides?

Technology Use

1. Which tools do you use for preparing and delivering oral presentations?
2. How could these tools be improved to better support dynamic presentations?
3. How could these tools be improved to better support QA sessions?
E.3 Protocol for Question Gathering Study

1. Thank you for coming in today to help us out with our research study. Let me first just double check that you are eligible. Are you at least 18 years old? Able to speak and read English? Have some college education? Have experience giving presentations? [If no, thank and dismiss, if yes continue.]

2. Let me tell you a little bit about what we will be doing in this study. We are developing a new technology that helps people during the question answering sessions after the oral presentations. To evaluate our support system, we want to understand what types of questions the audience asks during the question answering sessions. In this study, we will ask you to watch two short videos of me delivering oral presentations. After each viewing, I will ask you to come up with some questions regarding the presentation content that you would like to ask me after the presentation. In the end of the study, I will interview you about your own experience in question answering sessions. The study takes approximately 60 minutes, and you will be compensated $15 at the end of the session.

3. Does this sound like something you are still interested in? [If yes, continue, if no, dismiss]

4. Ok. Great. Now the first thing we will do is going over a consent form. I will explain it to you and you can take as much time as you want to read it and ask any questions you may have.

5. Now, I would like you to fill out a short questionnaire providing some information about yourself. Let me know when you are finished. [Hand participant socio-demographic]

6. The next questionnaire is going to ask you some questions about how you perceive your communication competence. [Self-perceived Communication Competence Scale]

7. Now we can get started. I would like you to watch a video of an oral presentation. In this video, I give a 3-minute presentation on [Nutrition/Exercise]. The goal is to talk about the importance of [Nutrition/Exercise], and to convince the audience to have a healthier lifestyle. I would like you to come up with at least 10 questions regarding the presentation content that you would like to ask me. You can take notes while watching the video. After you viewed the presentation, you will have 5 minutes to prepare your questions and then I will be here to answer your questions like a question answering session. [Play the 1st presentation]
8. [After viewing 1st presentation] Now, you will have 5 minutes to prepare at least 10 questions regarding the presentation.

9. [After 5 minutes, load the slides for the presentation] Now, I’ll be happy to answer your questions about the presentation. Like a question answering session, you can ask me to go to a specific slide. Would you mind if I record the questions?

10. [Second presentation] I would like you to view a second presentation. In this presentation, I give a 3-minute presentation, but on the topic of [Nutrition/Exercise]. [Repeat steps 7 - 9]
E.4 Protocol for Evaluation Study

[Before participant comes]

a. Make sure the Notes file has been generated beforehand
b. Make sure camera is charged

1. Thank you for coming in today to help us out with our research study. Let me first just double check that you are eligible. Are you at least 18 years old? And able to speak and read English? Do you have college education? Do you have experience giving presentations? [If no, thank and dismiss, if yes continue.]

2. Let me tell you a little bit about what we will be doing in this study. We are developing a new technology that helps people answer questions from the audience during the question answering session after an oral presentation. We call it the QA session for question answering. In this session, we will give you a pre-made PowerPoint slide deck and notes, and give you time to prepare. After your presentation, we will ask you to answer some questions regarding the content of your presentation with the help from different QA support systems. Your presentation and the question answering session will be videotaped. The session should take about 90 minutes, and you will be compensated $25 at the end of the session.

3. Does this sound like something you are still interested in? [If yes, continue, if no, dismiss]

4. Ok. Great. Now the first thing we will do is going over a consent form. I will explain the form to you and you can take as much time as you want to read it and ask any questions you may have. [Administer informed consent]

5. Now, I would like you to fill out a short questionnaire providing some information about yourself. Let me know when you are finished. [Hand participant socio-demographic]

6. The next questionnaire is going to ask you some questions about how you perceive your communication competence. [Self-perceived Communication Competence Scale]

7. Great, thank you. The next question is going to ask you some questions about your feelings of confidence as a speaker. These questions ask about your feelings associated with your most recent presentation [Personal Report of Confidence as a Speaker]

8. Ok, now we can get started. I would like you to give a short presentation on healthy lifestyle. Your audience will be a group of students and your goal is to convince them to have a healthier lifestyle.
We will break your presentation into 2 parts. You will prepare, and then deliver each part separately. Each part should be 3-minute long. After you present each part, we will have a 5-minute QA session when we will ask you to answer some questions related to the content of your presentation, using different QA support systems.

9. Now let’s start with the first part of your presentation. In this part, you will talk about [Nutrition/Exercise]. We have already prepared the PowerPoint slides and speaking notes for this part. There are 22 slides in total. But due to the time limit of 3 minutes, you will only present the first 7 slides. You can use the remaining slides and notes during the QA session to answer questions, if you want.

10. [Open the presentation slides] Now let’s look at the slides for the this part, about [Nutrition/ Exercise]. You will present the first 7 slides (until the Thank You slide). The rest are backup slides that you can display during the QA session if appropriate. Please do not change the slides.

11. [Show participants the notes] You can see the notes for each slide in this panel [show Notes pane for Slide 1]. The notes are already divided into main points. If you click on the title of each point, you will see more details about it. Please do not change the notes [Prompt participant to try clicking on a point]. You will be able to see these notes when you practice and deliver your talk, and during the QA session.

Some of the notes in the main slides are colored in red. These notes are backup content, which you don’t need to present during your 3-minute presentation. But we might ask you questions about this backup content during the QA session.

12. Great. Now, we I will give you 20 minutes to prepare for the presentation, including reviewing the content and rehearsing.
   a. First, please spend some time to review the slides and notes. Please review all the 22 slides, including the backup slides, because in the QA session we might ask you questions about the content of any of these slides.
   b. After you’ve reviewed the slides and notes, please rehearse the first 7 slides that you are going to present. During your rehearsal, we recommend you to speak aloud, as if you are presenting in front of an audience, but you can rehearse in any way you want.
   c. To start the rehearsal, just press the ‘Start Presentation’ button [Ask them to click on the button].
   d. Here is the remote control for you to advance the slides while you present. You can use the Right Arrow button to go to next slide, and Left Arrow button to go to previous slide. [Ask them to go to the first slide]
   e. You will see all the notes displayed on the notes pane here [point to the notes].
   f. Extra notes which you do not need to present will be displayed in yellow.
   g. After you finish your rehearsal, I will come in to show you how to use the QA system. And we can practice with a sample question, before you deliver your actual presentation.
   h. [Exit the presentation mode] OK, now it’s time to review the content and rehearse. I will come in after 20 minutes to show you the QA system. If you are ready before that, just let me know – I will just be waiting outside.
   i. [If participant asks for papers to take notes] Part of our study is to see how technology can support people who don't take notes on paper, so we would prefer you not using paper notes.
   j. [If participant asks for more time] Unfortunately, as part of the study we need to stay on schedule, so we're afraid we won't be able to give you more time. But you still have time now to rehearse.
18. [After Rehearsal, practice QA] Now let me show you how the QA system works. As I mentioned earlier, after your talk, I will ask you questions about the content of the presentation.
   a. For each question, I would like you to repeat the question in your own words and then try to answer the question. While answering the questions, you can display the related slides if you want. You can use any of the 22 slides, including the backup slides.
   b. The slide navigation bar at the bottom of the screen shows you all the slides in the presentation [point to the navigation bar]. You can scroll through the slides using the scrollbar at the bottom. [point to the scroller].
   c. When you hover the mouse over the slide thumbnails in the navigation bar, you can see their content, but they will not be shown to the audience yet. If you want to show a slide to the audience, just click on the corresponding slide thumbnail. The numbers above thumbnails show the slide number.
   d. [Control Condition]: The slides are ordered based on their order in slides deck.
   e. [Dynamic Condition]: After each question, the system uses this microphone to listen to you when you repeat the question in your own words. Based on your speech, it will automatically detect what slides are more likely to contain the answer to the question. The navigation bar will be updated dynamically to show the top 5 most relevant slides, and you can scroll to see the rest of slides using the scrollbar. Above each slide thumbnail, you will see a green bar. The size of the green bar shows how likely that slide includes the answer.
   f. [Dynamic Condition]: If you click on a thumbnail to show the slide to the audience, the system will also highlight the note topic that is most relevant to the question. You can highlight other topics by clicking on the up or down buttons.

19. Now let’s practice using the QA system. I will ask you a sample question. I would like you to repeat the question in your own words, then use the system to answer my question. Don’t worry if you cannot find the answer for the question, just let me know. Let me know when you are ready to practice [Practice QA]

20. [End of rehearsal, before delivery] Now it is time for you to deliver your presentation. But before you start your presentation, I just need you to fill in a questionnaire about how you feel right now [State Anxiety]

21. Great. As I mentioned before, we will videotape your presentation and the QA session. If you want to introduce yourself in the presentation, please introduce yourself by your first name only. For confidentiality, we don’t want your full name to appear in the presentation video. After your presentation, we will start the QA session and I will ask you 8 questions. If you don’t know the answer for any of the questions, just let me know and we can move on to other questions.

22. Whenever you are ready, please press the “Start presentation” button. [Press Record button]

23. [After presentation read the 8 questions chosen for the participant and wait for them to reply to the question. If the participant doesn’t understand the question you may repeat parts or all of the question]

24. Great. Now I would like you to fill in a questionnaire about your experience with the QA support system that you used during your presentation [Question Answering System Rating]

25. [Second condition: Open the Second Presentation] Now let’s move on to the second part of the presentation. In this part, you will be talking about [Nutrition/Exercise] [Repeat steps 9-19]
26. [After all conditions] Now I would like to ask you to rank the two question answering support systems that you have used. In this questionnaire, please rank the support systems based on each of these criteria, 1st (Better) or 2nd (Worse). [Ranking questionnaire]

27. That concludes the computer portion of the study. Lastly, I would just like to ask you a few questions about your experience with the tasks. [Conduct semi-structured interview, use phone to record, refer to interview sheet]
E.5 Interview Guide for Evaluation Study

A. General presentation experience (keep this short)

1. How often do you give presentations?
   a. What types of presentations?

2. How do you prepare for your presentation?
   a. Do you rehearse by speaking aloud?

3. Do you prepare backup materials, such as extra slides?
   a. If yes, how do you organize those materials? How useful are they to you?
   b. If using extra slides, how big is your extra slide set compared to your main slide set?

4. Do you write notes for your presentations?
   a. What types of notes do you prepare (e.g. full script, bullet points, keywords)?

5. How do you prepare for QA sessions?
   a. What are the main challenges you face during QA sessions?
   b. How often do you use slides to help you answer audience’s questions?
   c. If yes, do you have any problems accessing and navigating to relevant slides?

B. Presentation and QA system feedback

6. How do you feel about the presentations that you gave today?
   a. In terms of the presentation of the main slides, which presentation do you think you did better, and why? Which presentation do you think needs the most improvement, and why?

7. In terms of the QA session, how do you feel about answering the questions after the presentations? Which one do you think you did better in answering questions? Why?

7. Let’s talk about the QA support systems. How do you feel about the [Dynamic/Static] system?
   a. How do you think the system helped you? (in terms of easiness to find the answer, reliability, confidence boost)
   b. What did you like least about the system?
   c. What did you like most about the system?

8. Between the Dynamic and the static QA systems, which one do you prefer? Why?

9. [Dynamic condition] How accurate was the system in terms of detecting what slides and topics contain the answer?
   a. How useful was the likelihood indicator bar? How accurate?
   b. How did you feel when the slide navigation bar dynamically updates to display the most relevant slides?
   c. How useful was the topic highlighting? How accurate?

10. Would you use any of these QA support systems in the future? Why/why not?
a. Are there other features that you would like to add to these systems?

b. What are other situations that you think the systems might be helpful for (e.g. in meetings?)
# Appendix F

## Questionnaires for Quester Study

### F.1 Question Answering (QA) Rating Questionnaire

Please answer the following questions about the QA system:

*Write an ‘X’ on each line (on one of the dots):*

<table>
<thead>
<tr>
<th>Question</th>
<th>Rating Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>How easy was it to use the QA system?</td>
<td>Very difficult, Not at all, Very easy</td>
</tr>
<tr>
<td>How reliable was the QA system?</td>
<td>Not at all, Very much</td>
</tr>
<tr>
<td>How much do you feel the QA system helped you?</td>
<td>Not at all, Very much</td>
</tr>
<tr>
<td>How distracting was the QA system?</td>
<td>Very distracting, Not at all</td>
</tr>
<tr>
<td>How satisfied are you with the QA system?</td>
<td>Not at all, Very satisfied</td>
</tr>
<tr>
<td>How much would you like to give future presentations with the system?</td>
<td>Not at all, Very much</td>
</tr>
<tr>
<td>With the QA system, it was always obvious what the answer is.</td>
<td>Disagree, Agree</td>
</tr>
</tbody>
</table>

With the QA system, it was always obvious what the answer is.
F.2 Presentation Self-Rating Questionnaire

Please answer the following questions about the presentation you just delivered:

Write an ‘X’ on each line (on one of the dots):

How nervous were you during your presentation?

<table>
<thead>
<tr>
<th>Very poor</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very much</th>
</tr>
</thead>
</table>

How competent were you during your presentation?

<table>
<thead>
<tr>
<th>Very poor</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very much</th>
</tr>
</thead>
</table>

How would you rate the overall quality of your presentation?

<table>
<thead>
<tr>
<th>Very poor</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>•</th>
<th>Very good</th>
</tr>
</thead>
</table>
F.3 Ranking of the Teleprompter Systems

*Please indicate which of the 2 QA systems is better based on each criterion:*

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Ranking</th>
<th>Which system is better?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of use of the system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability of the system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helpfulness of the system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The level of distraction caused by the system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The system that made it obvious what the answer is</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with the system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desire to continue using the system</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix G

Slides Used in Quester Study

G.1 Nutrition and Diet

Slide 1

Nutrition and Diet
Nutritional Science is the study of the effects of food on the health, performance and disease resistance.

Our diet is based on the availability, processing and pleasantness of foods. A poor diet may cause health problems.

Nutrients or food components are two types: macro-nutrients which are needed in large amounts, and micro-nutrients which are needed in smaller amounts.
Carbohydrates are a common source of energy. They are a large part of foods such as rice, noodles, bread.

Proteins are one of the building blocks of body tissue. Sources of protein include meats, dairy products, grains and nuts.

Fibers are another essential macronutrient that can be found in whole grains, fruits, and vegetables.

Fats and water are other macronutrients
Dietary minerals are inorganic chemical elements that our body requires. Minerals are often artificially added to the diet as supplements.

Some of recommended minerals are Calcium, Chlorine, Magnesium, and Sodium.

Some vitamins are necessary for good health. Vitamin deficiencies may threaten our immune system and mental health.

13 types of vitamins currently recognized:

A (liver, orange, yellow fruit, orange vegetables, spinach)
B1 (Pork, oatmeal, brown rice, live, eggs, vegetables, potatoes)
B2 (dairy products)
B3 (also known as vitamin PP) (meat, fish, mushrooms)
B5 (meat, broccoli, avocados)
B6 (meat, vegetables
B7 (leafy vegetables, liver, peanuts)
B9 (pasta, bread, cereal, leafy vegetables)
B12 (meat, fish, eggs, milk)
C (fruits, veg, liver)
D (fish, eggs, liver, mushrooms)
E (fruit, veg, nuts, seeds)
K (leafy green vegetables, egg yolks, liver)
Malnutrition means insufficient, excessive, or imbalanced consumption of nutrients.

The recent increased consumption of sugar has caused the rise of some disorders such as diabetes, obesity, and even heart disease.

Processed foods generally have less nutritional value compared to whole, fresh foods and often contain potentially harmful substances.

Graph shows the percentage of people who get enough, not enough, and too much of certain food groups.
- orange bar shows the numbers of people who get less than the recommended amount and the blue shows % people who get at least the recommended amount.
The World Health Organization recommendations:

Eat roughly the same amount of calories that your body is using.

Limit intake of fats, sugar and salt.

Increase consumption of plant foods, particularly fruits, vegetables, whole grains and nuts.

recommend total water intakes of around 2 liters per day for adults.

Also, the U.S. Department of Health provides a sample week-long menu that fulfills the nutritional recommendations.

2 liters of water = 67.6 fluid ounces = 8.5 cups

Daily: Bread/ pasta/ rice 6-11 servings; vegetables 3-5 serv; fruits 2-4 serv; Dairy 2-3 serv; meat/eggs/ beans/ nuts 2-3 serv; very little oil or sugar
Thank You!
Reduction of carbs to 30% of diet and increase of fats to 40% of diet and increase of protein to 30%

- reduces hunger,
- causes weight-loss,
- decreases chances of getting cancer,

improves digestion and hormone regulation

massive reduction in carb to 10% or less of diet can result in fatigue/ irritability, known as “carb-flu”
Pork is high in Leucine which is a muscle protein. Chicken, turkey, eggs and seafood are good sources of lean protein.

Red meat is the most acidic protein source - therefore, best for addressing protein deficiency. Too much animal protein can increase the risk of cancer and kidney disease.

Lack of protein causes joint pain, unstable blood sugar levels, low energy, unable to sleep through night.
Carbs provide energy
Proteins build and repair tissue (skin, muscle) and hormones and enzymes
Fats store energy, insulate body, protect organs work with proteins
Plant Foods

- Vitamins
  - Some vitamins are only found in plant foods (e.g. vitamins C and E)

- Fiber
  - Found in plant foods
  - Helps weight loss
  - Fights high blood sugar

- Plant Protein

Helps weight loss, lower rates of cancer, lower cholesterol, lower blood pressure, reduced risk of diabetes

*Kale* is best plant-food. Contains large quantities of: Vitamin C, Vitamin A, Vitamin K, Potassium, Vit B6, Calcium, magnesium, copper, fibre and protein AND Cancer-fighting compounds

Plant food is a better source of fiber than fiber supplements. no evidence that fiber supplements have all the health benefits from food

Frozen vegetables generally have the same nutrition value as fresh vegetables, though nutrients like vitamins C and B are water-soluble and therefore will sometimes decrease in the freezing process.

Plant foods: have more fiber, and less fat and a plant-based diet can increase life-expectancy. Plant protein is safer than some red-meat-based sources of protein but animal protein does contain a greater variety of essential amino acids
Minerals

- Minerals are inorganic nutrients
- Not carbon-based
- Essential for the life of the human body
- Some Important Minerals for the body:
  - Calcium (good for bones)
  - Magnesium (important for kidneys and heart)
  - Phosphorous (good for bones, teeth and cell-functions)
  - Iron (essential for red blood cells)
  - Zinc (wound-healing)

10 essential minerals and their sources:
- Calcium (Dairy, dark green veggies, fish)
- Magnesium (whole grains, beans, nuts, cereals seeds)
- Phosphorous (dairy, meat, mushrooms, eggs, legumes)
- Iron (meat, fish, liver, lentils, apricots)
- Zinc (kelp, soybeans, black-eyed peas, lima beans, pumpkin, eggs)
- Sodium (kelp beats, okra, coconuts, unrefined sea-salt, dried fruits: raisins, figs, apricots)
- Selenium (seafood, organ meats, dairy, butter, sesame seeds, garlic, brazil nuts)
- Potassium (bananas, potatoes, tomatoes, avocados, citrus fruits, spinach)
- Iodine (kelp, salt, swiss chard, spinach, turnip, squash)
- Manganese (leafy greens, avocados, blueberries, pineapples, kelp, green tea, beetroot, eggs)

Inorganic nutrients: nutrients which are **not carbon-based** (as opposed to proteins, carbohydrates, sugars, vitamins)
Excess vitamin intake and Vitamin deficiency can both have a variety of effects:
- Developmental issues in children
- Increased risk of skin, kidney and heart disease
- Increased risk of bone fractures and bone weakness

Excess vitamin intake generally caused by taking supplements and fortified food, not by eating ordinary food alone:

Effects of excessive intake of...
- vitamin A: skeletal problems (in children), higher risk of osteoporosis and hip fractures (in adults)
- vitamin B3 (niacin): flushing reactions, tingling, itching and reddening of the skin, rashes
- vitamin C: nausea, diarrhea, stomach cramps
- vitamin D: nausea, weakness, kidney problems

Effects of deficiency of...
- vitamin A: immune system dysfunction and weakness, problems with eyesight and reproductive function
- vitamin B3: skin disease, digestive tract problems
- vitamin C: Scurvy! (which involves easy bruising and bleeding, joint and muscle pain)
- vitamin D: bone fractures, muscle weakness, difficulty thinking clearly, fatigue
Artificial vs. Natural

➢ What about Supplements?
  ➢ Contain specific vitamins and minerals
  ➢ Less complex and varied in nutrition content than food
  ➢ Can address specific deficiencies
  ➢ Cannot make up for nutrition content of food

➢ What about GMO?
  ➢ Stands for “genetically modified
  ➢ Mixed views on health risks:
    ➢ Some studies point to increase in certain diseases among populations working with GMO products
    ➢ Link remains unclear

Artificial supplements tend to be less nutritious and good for your body than food, which generally contains more nutrients in more complex structures than supplement pills.

Nutrient can help people with specific nutrient deficiencies but cannot adequately provide all of the nutrients which food provides.

Some research suggests exposure to genetically modified products increases risk of disease

  e.g.: In 2003, approximately 100 people living next to a Bt cornfield in the Philippines developed skin, respiratory, intestinal reactions and other symptoms while the corn was shedding pollen.
Malnutrition compromises immune system (increasing risk of other diseases) and development problems;
cognitive effects include attention deficit disorder, impaired school performance, decreased IQ scores, memory deficiency, learning disabilities, reduced social skills, reduced language development, reduced problem-solving abilities
obesity, heart disease, kidney disease, increased risk of cancer, fatigue, decrease life-expectancy
Malnutrition is a bigger problem the world over than obesity - people with obesity can and often do have deficiencies of certain nutrients even though they have excess intake of fats and carbs.
Processed Food

Common Processed foods:
- breakfast cereals
- cheese
- canned vegetables
- bread
- meat products, such as bacon, sausage, ham, salami and paté
- ‘convenience foods’, such as microwave meals or ready meals
- cakes and biscuits
- Drinks, such as milk or soft drinks

Problems include:
- High salt content
- High Sugar Content
- High Fat content

High salt, high sugar and high fat increase: risk of heart disease, cancer and other major diseases.

Diet soda drinkers still have increased risk of weight gain over non-soda drinkers; also higher chance of type 2 diabetes development
Fat, Sugar, Salt

- Fat
  - Insulates organs and aids work of proteins
  - In excess, can damage arteries and increases risk of heart disease and cancer
- Sugar
  - Provides energy
  - In excess, increase risk of heart disease and heart attack
- Salt
  - Regulates bodily functions
  - Increases blood pressure and risk of heart disease

Fat protects and insulates organs; aids proteins and enzymes in essential bodily repair work.

Excess causes heart disease; arteries wall thickening which causes stroke and heart attack.

Fat intake should be around 0.4 to 0.5 grams per pound of your target body weight.

Increases risk of heart disease and heart attack. sugar does not on its own lead to eating disorders, but can lead to diabetes

Increases blood pressure: 1500 to 2300 mg per day (=0.75-3.75 teaspoons)
Calories are a unit of measurement for energy. Too many calories consumed (and not used up in activity) lead to weight gain; too few calories leave the body tired, weak, unable to function.

Recommended daily:

1500 to 2000 cal for average woman

2000 to 2500 cal for average man
Water (on its own, or contained in food or other fluids) is essential part of daily diet; cannot be supplemented by anything else

Daily amount *Including water from food*: 15 cups for average male; 11 for average female. also depends on body size and climate. “In rare cases, drinking an extreme amount in a short time can be dangerous. It can cause the level of salt, or sodium, in your blood to drop too low”

Cancer risk higher among drinkers of chlorinated water;
Fluoride in water:
- at low levels strengthens teeth
- at high levels can cause fluorosis, a discoloration of teeth
- at very high levels can cause skeletal damage
This is a menu aimed at weight loss, and incorporating daily exercise.

<table>
<thead>
<tr>
<th></th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
<th>Sunday</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Breakfast</strong></td>
<td>Egg scramble with tomatoes &amp; avocados</td>
<td>Protein shake w/ strawberries</td>
<td>Smoked salmon w/ bacon, seeded spinach, tomatoes</td>
<td>Protein shake w/ strawberries</td>
<td>Smoked salmon w/ mustard &amp; tomatoes</td>
<td>Egg scramble with tomatoes &amp; avocados</td>
<td></td>
</tr>
<tr>
<td><strong>Lunch</strong></td>
<td>Cucumber slices w/ ranch dressing</td>
<td>Cherry tomatoes</td>
<td>Cucumber slices w/ ranch dressing</td>
<td>Bell pepper slices w/ dressing</td>
<td>Cherry tomatoes</td>
<td>Bell pepper slices w/ dressing</td>
<td></td>
</tr>
<tr>
<td><strong>Dinner</strong></td>
<td>Chicken, steamed broccoli, chopped spinach salad</td>
<td>Grilled salmon w/ Mediterranean salad</td>
<td>Chicken breast, roasted broccoli, cauliflower &amp; sweet potato fries, mixed greens</td>
<td>Chicken breast, roasted broccoli, cauliflower &amp; sweet potato fries, mixed greens</td>
<td>Beef &amp; broccoli, mixed greens &amp; tomato basil</td>
<td>Beef &amp; broccoli, mixed greens &amp; tomato basil</td>
<td></td>
</tr>
<tr>
<td><strong>Snack</strong></td>
<td>Cherry tomatoes</td>
<td>Cucumber slices w/ ranch dressing</td>
<td>Cherry tomatoes</td>
<td>Bell pepper slices w/ dressing</td>
<td>Cherry tomatoes</td>
<td>Bell pepper slices w/ dressing</td>
<td></td>
</tr>
</tbody>
</table>

Projects in the field of “Nutrition literacy” aim to provide parents with the resources to provide children not just with healthy food, but with the knowledge and means to identify and understand health and nutrition information, from a young age.

“The new rules, passed by Congress in 2010 to fight childhood obesity and promote health, took effect in August. High school cafeterias must serve twice as many fruits and vegetables as before, limit proteins and carbohydrates, and serve lunches containing 750 to 850 calories.”
Information Resources

- Resources aimed at individuals, parents and teachers:
  - “Nutrition Literacy.” Diet.com
    https://www.diet.com/g/nutrition-literacy/get=nutrition-literacy#G
  - USDA
  - Nourish Interactive
    http://www.nourishinteractive.com/nutrition-education-printables/category/48-teacher-resources-teaching-kids-healthy-habits-nutrition-manuals

Three websites aimed at teachers, parents and individuals concerned with nutrition literacy and food education
G.2 Questions regarding Nutrition and Diet Slides

1. Does sugar lead to eating disorders?
2. 2 liters of water should be had each day; How many ounces would that be? And how many cups?
3. Should we have fluoride or chlorine in our water?
4. Are Genetically modified products healthy?
5. What are the features of malnutrition?
6. Are frozen vegetables healthy? Do they have the same amount of fiber as fresh vegetables?
7. Is ham a kind of processed food?
8. Beef, pork or chicken? Which is the best source of protein?
9. Is there any drawback to intake too much protein?
10. Some people try to get rid of fat is this healthy?
11. Is it healthy to have a low carb diet?
12. Is it healthy for a person to go for fat and protein only diet instead of carbs?
13. Are there any artificial products that provide fiber and If so are they better than natural sources?
14. Why is it bad to have processed food like diet soda if they have lower calorie?
15. What should be the appropriate salt intake of a person?
16. Are there negative consequences of excessive vitamins intake?
17. Is there something as too much water?
18. People just take supplements of nutrition What are the bad effects?
19. We have insufficient food in Africa but US has obesity what is the major problem to tackle today?
20. What is role of school to improve awareness on nutrition How good are school lunches?
21. I was curious to see the sample weekly menu?
22. Pyramid structure you showed, what does it say?
23. Where can I get more information about nutrition?
24. What would happen if I don't consume enough nutrients?
25. I want more explanation on the bar graph.
26. How harmful is processed food and what kind of harm?
27. Why do we need the same exact amount of calorie as the body needs?
28. What are natural ways to get minerals?
29. What are inorganic chemicals?
30. Why recommend more plant foods?
31. How do you decide the correct amount of calorie intake?
32. What are the most nutritious plant foods?
33. What are most common problems due to lack of protein?
34. How much fat should we consume in a day?
35. What are the different types of vitamins?
36. What are the problems associated with lack of each type of vitamin?
37. You should consume 2 liters of water; does it change based on climate condition?
38. Difference between plant nutrition and animal?
39. Why we need to limit amount of fat?
40. Can we eat a lot of fruit but not 2 liters of water?
G3. Physical Exercise

Slide 1
Physical exercise is any bodily activity that improves or maintains physical fitness and overall health and wellness.

It is performed for various reasons, including weight loss or maintenance, mental and physical fitness, or just enjoyment.

Physical exercises are generally grouped into three types, depending on the overall effect they have on the human body: Aerobic, anaerobic, and flexibility exercises.
Aerobic exercise is any physical activity that uses large muscle groups and causes the body to use more oxygen than it would while resting.

Examples of aerobic exercise include cycling, swimming, brisk walking, and hiking.

Anaerobic exercise are short and intense activities. They can strengthen, and tone muscles.

Examples of Anaerobic moves are push-ups, pull-ups, and bicep curls using dumbbells.

Flexibility exercises stretch and lengthen muscles. The goal is to improve the range of motion which can reduce the chance of injury.
Physical exercise is important for maintaining physical fitness. People who participate in moderate to high levels of physical exercise have a lower mortality rate compared to individuals who are not physically active.

Exercise reduces blood pressure, total cholesterol, and body weight. Physical inactivity is a risk factor for the development of cardiovascular diseases.

Moderate exercise benefits the human immune system.

A study published in JAMA showed that 75 minutes of exercise reduces mortality rate by 20%.

Effects of excessive exercise are not clear.
Exercise can improve academic performance and adult productivity. It can preserve cognitive function in old age.

Physical exercise can prevent depression and also be a therapy for depressive disorders.

Exercise generally improves sleep for most people, and helps cure sleep disorders such as insomnia.

Study by national sleep foundation shows that half of people with no activity get fairly to very bad sleep

Image shows two MRI brain scans of brain activity when sitting and active. Red indicate brain activity and the blue indicates brain inactivity. Walking increases brain activity
In the United States, public health officials recommend that every adult should do moderate exercise for a minimum of 30 minutes daily.

Nowadays, unfortunately, there has been a shift towards less physical activity.

Parents can promote physical activity and limit the amount of time children spend in front of screens.

Implementing physical exercise in the school system is essential.

Moderate exercise such as brisk walking or swimming
Thank you!
Builds stamina, helps with weight loss, improves respiratory and heart health, improves metabolism, relieves stress.

Muscle and joint pain, overexertion. Get a physical exam to know your health before exercising.
Anaerobic Exercise

- Develops Strength
- High Intensity
- Short Duration

CDC notes that strength training done three times per week reduces risk of heart disease, improves bone strength, elevates calorie burning, and improves mental health.

Soreness, Risk of injury, overtraining. Get medical exam beforehand. Also a personal trainer is recommended to plan routines.
To increase flexibility, you must stretch periodically during the day and especially after a workout. Stretching warms up the muscles and prepares the body for physical exertion.

Common types: Dynamic and Static stretching. Dynamic stretching improves range of motion. Static stretching is most common and is considered safe and effective for improving overall flexibility.
Each category has its own benefits.

A balance between exercise types is recommended.
Cardio or aerobic exercise is best for weight loss and weight management. However, it is only one part of the weight loss plan. Diet is equally important.

For building muscle, weight training and diet are just as important. A metabolism-boosting workout builds strength and tones your muscles.
Cardiovascular health refers to the health of the heart and blood vessel system. It is commonly affected by cardiovascular diseases that include diseased vessels, structural problems, and blood clots.

Regular exercise lowers risks of cardiovascular diseases. Exercising for 30 minutes or more on most days can help you lose weight, improve your cholesterol, and even lower your blood pressure.
The immune system is a network of cells, tissues, and organs that work together to defend the body against attacks by bacteria, parasites, and fungi that can cause infections.

Physical activity may help flush bacteria out of the lungs and airways, reducing the chance of getting a cold, flu, or other illness. Exercise causes changes in antibodies and white blood cells that fight diseases.
Positive link. It increases heart rate, which pumps more oxygen to the brain. It also aids the bodily release of hormones which nourish brain cells, improve memory function, enhance mood.

Research shows that exercise helps release neurotrophins which assure the survival of neurons in areas responsible for learning, memory, and higher thinking.
Exercise can treat mild to moderate depression as effectively as antidepressant medication. Exercise is a natural and effective anti-anxiety treatment, and also helps relax muscles and relieve tension in the body.

Short bursts of exercise in the morning or afternoon can help regulate sleep patterns. Relaxing exercises such as yoga or gentle stretching can help promote sleep.
General Recommendations

- Drink eight 8-ounce glasses of water a day
- Consume 2,000 calories per day.
- Balanced nutrition.

Water regulates your body temperature and lubricates your joints. It helps transport nutrients to give you energy and keep you healthy. If you're not hydrated, your body can't perform at its highest level.

Eat nutrient dense foods such as whole grains, lean proteins, fruits, and vegetables. Exercise results are greatly affected by eating habits.
Body types and mass differ not only for men and women but also people of the same sex. Endurance capacities are also different. Individual exercise plans are required.

Age, weight, physical disabilities also affect exercise. The average 30 minutes of physical activity needs to be altered depending on these.
Negative health effects develop when exercise is taken to extremes. Overexertion is to strain or put too much pressure on one’s self. Example: to run 10 miles when the body is only used to running five.

Overtraining can lead to an increased resting heart rate, muscle tear, insomnia, and in serious cases cardiac arrest.
Regular physical activity helps children build healthy bodies and prevent chronic disease. Aerobic activity should make up most of your child's 60 or more minutes of physical activity each day.

Running, skateboarding, dancing, playing basketball, swimming and gymnastics are examples kids’ exercise.
Set realistic expectations and establish routines that are suited to you.

Positive attitude, scheduled workouts, goal setting, group workouts and buddy system, personal trainer.
Information Resources

- https://www.health.harvard.edu/newsletter_article/exercise-resources
G4. Questions Regarding Physical Exercise Slides

1. How many repetitions of weights should we do?
2. What is the way that exercise prevents cardiovascular disease?
3. How much exercise should the children have each day?
4. How much stretching should we do each day?
5. How does exercise help the immune system?
6. Will exercise prevent dementia?
7. Should there be different exercises for girls and boys?
8. What about drinking water after or during exercise how much should we have?
9. At what age can kids start doing intense exercise?
10. Which exercise has a larger chance to cause injuries?
11. What is the critical level for over exercise Causing damage to body?
12. What is the mechanism beyond exercise helping to sleep?
13. In the plot you showed risk reduction with exercise first there is increase and then decrease what is the reason for the decrease
14. How does exercise affect the cognitive functions?
15. Which of the 3 categories of exercise is most effective?
16. What are the adverse effects of excessive and rigorous exercise?
17. What are the types of flexibility exercise?
18. What are the benefits of flexibility exercise?
19. How does exercise affect depression?
20. Does the time and rigorousness of exercise have to be changed based on the overall health of a person?
21. How can we motivate and promote exercise among adults?
22. You said exercise helps in weight loss but a new study says it does not help in weight loss and food is important.

23. I didn't understand the brain scan image. What does it mean?

24. Which category of exercise is more recommended for a normal life or is it a blend of them?

25. Where can I get more info to follow up on this info?

26. Talking about the right type of food which supports exercise What if I don't eat well but exercise everyday?

27. Is exercising in particular hour of the day more efficient?

28. You said 30 minutes of moderate exercise what is moderate exercise?

29. Please explain more about the bar graph on sleep?

30. What are cardiovascular diseases is it just heart attack?

31. Exercise affects academic performance what is this based on?

32. You mentioned statistics on mortality and exercise what study is this based on?

33. You mentioned 30 minutes of exercise does it depend on age?

34. How many times a week should we do aerobic exercise?

35. How often you should do anaerobic exercises?

36. What are some of the exercises that help you with cardiovascular health?

37. What are risks of over doing aerobic exercises?

38. How can government and schools help in spreading awareness about benefits of exercise?

39. Should we do flexibility before or after other types of exercise?

40. If someone is too lean will exercise help in gaining weight?


18. Andrew Caines and Paula Buttery. The effect of disfluencies and learner errors on the parsing of spoken learner language. in.
26. Anne Dohrenwend. Serving up the feedback sandwich.


34. Yasuhsa Fujii, Norihide Kitaoka and Seiichi Nakagawa. 2007. Automatic extraction of cue phrases for important sentences in lecture speech and automatic lecture speech summarization. in Eighth Annual Conference of the International Speech Communication Association.


43. Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen and Tara N


68. Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard and David McClosky. 2014. The stanford corenlp natural language processing toolkit. in *ACL (System Demonstrations)*.

85. Erwan Pépiot. 2014. Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. in *Speech Prosody* 7, 305-309.


